Precision measurements of indirect *CP* violation in the charm sector with LHCb

A thesis submitted to The University of Manchester for the degree of

Doctor of Philosophy

(PhD)

from the work of

Kevin S. Maguire[†] **BSc.**

in the Faculty of Science and Engineering

School of Physics and Astronomy

Particle Physics Group

2018

[†]Supervised by Professor Christopher Parkes and Dr. Marco Gersabeck

Contents

Word count: 44,732

List of Figures

List of Tables

Abstract

A study of indirect CP violation in D^0 mesons is presented through a measurement of the asymmetry of the effective decay widths of D^0 and $\overline{D}^0$ mesons. This observable, known as A_Γ, is measured in decays to the CP eigenstates K^+K^- and $\pi^+\pi^-$ using data collected in pp collisions by the LHCb experiment in 2012 with an integrated luminosity of $1.9\,\mathrm{fb}^{-1}$. The results are combined with a previously published LHCb measurement using $1\,\mathrm{fb}^{-1}$ of data collected in 2011 to yield average values of,

$$\begin{aligned} A_\Gamma(K^+K^-) &= (-0.14 \pm 0.37 \pm 0.10) \times 10^{-3}, \\ A_\Gamma(\pi^+\pi^-) &= (\;\;\,0.14 \pm 0.63 \pm 0.15) \times 10^{-3}, \end{aligned}$$

where the first quoted uncertainty is statistical, and the second systematic. The results show no evidence of CP violation and improve on the precision of the previous best measurements by nearly a factor of two.

New methods are presented for measuring the D^0 meson lifetime that rely on updated LHCb data acquisition techniques, which collect large datasets while reducing time-dependent biases. Many measurements of indirect CP violation observables depend on precision measurements of meson lifetimes. Furthermore, the large datasets being collected by LHCb necessitate faster analysis techniques such that results can keep pace with the magnitude of acquired data. The methods presented employ a binned fit, where the main method separates an intractable background with a simple selection requirement and a second potential method investigates novel machine learning classification algorithms.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Kevin S. Maguire BSc.
University of Manchester
Faculty of Science and Engineering
School of Physics and Astronomy
Particle Physics Group
Oxford Road, Manchester, M13 9PL

Copyright

1. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the Copyright) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

2. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

3. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the Intellectual Property) and any reproductions of copyright works in the thesis, for example graphs and tables (Reproductions), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

4. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy [1], in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations [2] and in The Universitys policy on Presentation of Theses

[1] see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420
[2] see http://www.library.manchester.ac.uk/about/regulations/

Acknowledgements

First and foremost I would like to thank my parents for their constant support and guidance, in all the forms that it has taken. Most of all I thank them for the great courage and daring they have shown by changing their lives for the better in search of fulfillment. Their clear example that happiness is attainable and that purpose is imperative, motivates me to aim high and work hard towards my personal goals.

I thank Sarah for not allowing me to ruin our future when I first left Dublin. I thank her for perseverance over the two years that we were apart and for her willingness to move to Manchester in my third year. Living together in Manchester has been wonderful for me and the best decision that we have so far made together.

I thank also my supervisor Chris for believing that I was capable of this degree with frankly not much evidence given on my part. I was lucky to have first worked with Chris when I did a summer internship at CERN in 2013. His belief that I had the ability, and his influence, has given great value to me and my future career. I hope that he is satisfied with the work contained in this thesis and the work that I have done under his supervision. His patience and his great depth of knowledge in the field have been invaluable to me.

Thanks is also due to my co-supervisor Marco, especially in recent times when Chris has been laden with other responsibilities. I have appreciated his ability to point out where I have gone wrong and to help in fixing those problems.

I haven't often spoken to Mika since I started my PhD, but I must thank him for the introduction he gave me to particle physics and computing when I was a summer student. Mika was my closest contact at CERN during that work and I often think of the things he taught me when I am teaching other students about our computing systems. In this regard I would also like to thank Mark, Michael and all other PhD students and post doctoral researchers who helped me to understand the technical workings of the collaboration when I first began this degree.

This degree would not have been possible without support from The University Of Manchester and its president, Dame Nancy Rothwell, through the President's Doctoral Scholarship (PDS). Some funding was also provided by the Science and Technologies Facilities Council (STFC) of the UK government. I thank these organisations for their commitment to scientific endeavor and blue sky research. Congratulations also to the UK government who lured me to this country with funding opportunities, and where I may stay for some time yet.

I have enjoyed making many new friends while doing this degree. The value and comfort that you have brought to my life in the last few years has been essential. The personal and professional guidance that you have all provided has been of great value to me. Thanks in particular to my LHCb and ATLAS friends and colleagues at The University Of Manchester. I will remain with you in Manchester for some time yet.

The author

Education

- **The University of Manchester**,
  Doctor of Philosophy, PhD,
  Thesis title: "Precision measurements of indirect *CP* violation in the charm sector with LHCb",
  Supervised by Professor Christopher Parkes and Dr. Marco Gersabeck,
  2014 – 2017

- **European Organisation for Nuclear Research (CERN)**,
  Summer student, 2013

- **University College Dublin**,
  Bachelor of Science, BSc.,
  First class honours degree,
  Grade point average: 4.11/4.2
  2010 – 2014

- Newpark Comprehensive School

- Rathmichael National School

Professional Roles

- **API Cognition Ltd.**,
  Applied Cognition Engineer,
  Norwich, UK
  2017 – present

- IOP HEPP Subgroup,
  Experimental student representative,
  2015 – 2017

Part I

Introduction and setting

Chapter 1

Introduction

Particle physics is the study of the fundamental building blocks of matter and the interactions between them. The aims of such studies are, to understand the underlying physics of the universe and predict its behaviour, to inform that understanding with experiments, and further refine theoretical models to improve predictability.

One of the unanswered question of this field is the observed difference between matter and anti-matter in the universe. These two forms should be treated equally by nature, yet matter appears to be favoured over anti-matter, as shown by its relative abundance in the universe. The commonly proposed solution to this problem is the violation of a matter and anti-matter symmetry called *CP*. However, this cannot be the complete solution as a necessary condition to explain the matter and anti-matter imbalance is baryon number violation, which is not fulfilled by *CP* violation. A review of the theoretical framework of particle physics and how it relates to the work of this thesis is presented in Chapter 2.

This thesis is primarily concerned with experimental particle physics measurements which are motivated by theory. Presented here is a measurement of A_Γ, one of many interesting observables that is used to test *CP* conservation in fundamental physics reactions. This observable is the asymmetry between the lifetime of the D^0 meson and its anti-particle, the $\overline{D}^0$ meson. In measuring A_Γ, the individual lifetimes of the D^0 and $\overline{D}^0$ mesons are measured in their decays to two pions, or two kaons. While *CP* violation has already been observed in similar processes, those investigated in this work show no evidence of *CP* violation and thus are consistent with *CP* conservation. The small scale of observed violations and the lack of observed violation in other processes indicate that *CP* violation may not be sufficient to explain the matter and anti-matter asymmetry in the universe.

The data used in the measurement of A_Γ were collected by the LHCb experiment at the Large Hadron Collider (LHC). This is the worlds largest particle collider with a circumference of 26.7 km and is located under the French-Swiss border near Geneva. The data were collected in 2012, and the final measurement is combined with another measurement of A_Γ from data collected in 2011. The selection of the data sample used, the measurement method, and the resulting value of A_Γ are presented in Part II.

The measurement of A_Γ in this thesis was a success and represents one of the two best measurements ever made of this observable. However, there are a number of potential issues with continuing this type of measurement method on datasets collected by LHCb in more recent years. The method presented in Part II requires a large amount of computing time, which increases with the number of collision events in the dataset. It also requires the dedicated efforts of a number of analysts for an extended period of time. The datasets collected by LHCb from 2015 to 2017 are larger than were taken in 2011 and 2012, thus compounding these problems. Also the rate of data collection will increase again after the LHCb upgrade in 2019/2020. A final issue is that the method has underlying problems which may make it unsuitable for other *CP* symmetry measurements.

In Part III, potential new methods for measuring the D^0 meson lifetime are studied. These new methods are comparatively simpler, and thus take less computing time and less analyst time. As a demonstration, the lifetime of the D^0 meson is measured from data collected by LHCb in 2016.

Chapter 2

Theoretical setting

The modern theory behind particle physics is known as the "standard model" (SM). It is based on quantum field theory and group theory, and has been built off the centuries of work that came before, including the development of Maxwell's equations [1] and the experimental discovery of the positron [2]. The formulation of the SM as we know it today was developed in the latter half of the 20th century. In the early years of its formulation, new additions were driven by experimental observations of new particles. In more modern times the theory has made a wealth of predictions that have so far been proven correct in nature. One of these many successes is the bound on the electric dipole moment of the electron [3], which is an accurate test for potential additions to the SM, and has implications for *CP* violation.

Indeed the great success of the theory has been the biggest problem for physicists in recent years. The apparent completeness of the SM has resulted in experimental searches becoming routine consistency checks. As a consequence, abnormalities between predictions and experimental results have become much sought after and been given the reverent moniker of "new physics" (NP).

While there are great successes, there are also obvious problems. The SM is by no means a complete theory. For example, the neutrinos are predicted to have zero mass, while observations of their flavour oscillations prove that they do indeed have mass [4,5]. There are also a number of astronomical observations that highlight the incompleteness of the theory. These are the imbalance of matter and anti-matter in the universe, which is not quantitatively accounted for in the SM, and the necessity for dark matter in the cosmos, for which there is no convincing SM candidate.

This chapter contains an introduction to the SM as well as a primer on *CP* violation, which is integral to the studies performed in this thesis. In Section 2.1 the SM in its modern form is presented. Section 2.2 introduces the quantum mechanical operators of interest for the studies to follow, while Sections 2.3 and 2.4 present flavour mixing and the violation of the *CP* symmetry within the SM. Finally, Section 2.5 introduces the SM observable to be measured in this thesis.

2.1 Standard model of particle physics

The SM is constructed of two families of spin-1/2 "matter" particles called the "leptons" and the "quarks" as well as twelve "force carrier" particles, which are spin-1 states known also as "bosons". The bosons describe the three fundamental forces of nature, electromagnetism, weak nuclear and strong nuclear. Gravity is the fourth fundamental force, but is neglected in the SM, as its coupling strength to matter particles is many orders of magnitude smaller than the couplings of the other forces. Interactions between matter particles occur by the exchange of momentum via a carrier boson. A spin-0 particle called the Higgs boson is the final piece of the SM and is responsible for the mechanism that gives mass to the other particles. Most particles have an associated anti-particle, which shares some of the same properties but have opposite values of other properties. This comes as a consequence of negative energy solutions to quantised equations of motion. In exception to this are particles that are their own anti-particle, such as photons.

2.1.1 Matter content

The first of the groups that encompass the SM is that of the matter particles. A "fermion" is a spin-1/2 particle state with a wavefunction that satisfies the Dirac equation,

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0, \tag{2.1}$$

where summation over the repeated indices is implied and the four γ^{μ} matrices are the "Dirac matrices", and have the usual definition. A fermion is defined by the spin-statistics theorem, which says that the wave function of a system of identical spin-1/2 particles is anti-symmetric under the exchange of any two particles. A corollary of this is the Pauli exclusion principle, which says that no identical fermions can occupy the same quantum state. Or in other words, no two fermions in a single system can have the same quantum numbers. This principle had fundamental consequences for the development of the SM, as the

Table 2.1: Properties of the leptons, taken from [9]. Shown are, Q, the electric charge, I_z, the z component of weak isospin and L_i, the value of lepton flavour i. The electron and muon masses are not shown to their full precision.

Name	Symbol	Mass (MeV/c^2)	Spin	Q	I_z	L_e	L_μ	L_τ
Electron	e^-	0.511	1/2	-1	-1/2	+1	0	0
Muon	μ^-	105.66	1/2	-1	-1/2	0	+1	0
Tau	τ^-	1776.86 ± 0.12	1/2	-1	-1/2	0	0	+1
Electron Neutrino	ν_e	$< 2\,\mathrm{eV}/c^2$	1/2	0	1/2	+1	0	0
Muon Neutrino	ν_μ	< 0.19	1/2	0	1/2	0	+1	0
Tau Neutrino	ν_τ	< 18.2	1/2	0	1/2	0	0	+1

discoveries of new particles that appeared to violate the exclusion principle were clear evidence of new internal quantum properties.

The family of fermions known as the "leptons" are those elementary particles that do not interact via the strong nuclear force. As shall be seen, this is because the leptons do not carry "colour" charge, which is the conserved quantum number of the strong interaction. Electrically charged leptons interact via the weak nuclear and electromagnetic (EM) forces, while neutral leptons interact through the weak nuclear force only. As the weak force has a characteristically weak coupling to matter, the neutral leptons, or "neutrinos" rarely interact with other matter and thus are difficult to detect. The "flavour" of a lepton is a fundamental property that takes one of three values, electron, muon or tau, along with the three corresponding anti-flavours. These quantum numbers are thought to be conserved in all interactions. Recently, tension with the assumption of lepton flavour conservation has been seen in a number of experimental measurements [6–8]. Shown in Table 2.1 are the properties of the leptons.

From the late 1940s and throughout the following decades a large variety of particles were discovered that at that time were called "hadrons". A theory for describing this particle "zoo" was proposed by Gell-Mann in 1964 by introducing three new fermions labelled "quarks", as well as their three anti-particles [10]. The zoo of particles is described by states of these quarks bound by the strong nuclear force. These bound states were made from either one quark and one anti-quark, called "mesons" or three quarks, called "baryons". These initial three quarks were sufficient to describe the known hadrons, but as time went on, a

Table 2.2: Properties of the quarks, taken from [9]. Shown are, Q, the electric charge, I_z, the z component of weak isospin, S, the strangeness number, C, the charmness number, B, the bottomness number and T, the topness number. The quantum number $I_3 = \frac{1}{2}(N_u - N_d)$ is the third component of the quark "isospin", where N_i is the number of quarks of type i in a state.

Name	Symbol	Mass (MeV/c^2)	Spin	Q	I_z	I_3	S	C	B	T
Up	u	$2.2^{+0.6}_{-0.4}$	1/2	2/3	1/2	1/2	0	0	0	0
Down	d	$4.7^{+0.5}_{-0.4}$	1/2	-1/3	-1/2	-1/2	0	0	0	0
Charmed	c	1280 ± 30	1/2	2/3	1/2	0	0	1	0	0
Strange	s	96^{+8}_{-4}	1/2	-1/3	-1/2	0	-1	0	0	0
Top	t	$173{,}100 \pm 600$	1/2	2/3	1/2	0	0	0	0	1
Bottom	b	4180^{+40}_{-30}	1/2	-1/3	-1/2	0	0	0	-1	0

further three were introduced. Similar to the lepton flavours, the quarks also have associated flavours. The strong nuclear and EM forces conserve these flavours in their interactions with matter, along with the neutral weak nuclear current. In contrast, the charged weak nuclear current never conserves them in its interactions with quarks. Shown in Table 2.2 are the properties of the quarks.

Two problems arose with the quark model of hadrons. The first was due to the fact that symmetric baryon wavefunctions did not predict the observed lowest baryon states, even though the theory suggested that they were symmetric. Secondly, an Ω^- hadron was discovered, which is a bound state of three identical d quarks, seemingly in violation of the exclusion principle [11]. These problems were solved by Greenberg in 1964 by introducing a "colour" degree of freedom that is anti-symmetric under exchange, and introduces a new colour charge quantum number [12]. Thus the total wavefunction is anti-symmetric, and the three d quarks are not identical, but have different "colours". Each quark has one of three colour values, red green or blue, or the corresponding anti-colour values.

It is found experimentally that the weak nuclear force is maximally violating of parity [13]. The weak force bosons only couple to left-handed particles and right-handed anti-particles. The handedness, or "chirality", of a massless particle state is determined by its eigenvalue with respect to the combined spin and parity operator $S \cdot P$. For massive particles, the handedness depends on how the particle state behaves under the Lorentz transformation.

As right-handed neutrinos are colour and electrically neutral, they do not interact via any force and thus may not exist. The field for a fermion can be split into its left- and right-handed parts, using handedness projection operators.

2.1.2 Gauge theory

The SM is based on the product of three mathematical groups[1], $SU(3) \times SU(2) \times U(1)$ [14–16]. The EM force is parameterised as the unitarity group, the abelian (commuting) group of complex numbers with absolute value of one, $U(1)$. Unifying the weak nuclear and EM forces into the electroweak force results in an alternative representation of both forces, in which the weak force is described by the special unitary group of 2×2 matrices with determinant equal to one, $SU(2)$. The extra degree of freedom in the weak force results from the importance of chirality in the interactions. Finally the strong force is represented with yet more degrees of freedom in the special unitary group of 3×3 matrices, due to three colour charge values. In general, the $SU(N)$ group of $N \times N$ matrices is non-abelian, the consequences of which will be seen in the following section.

A gauge theory is a field theory in which the associated Lagrangian is invariant under some transformations. The group properties of these transformations define the "gauge group". A set of elements, $\{\dots, T_j, \dots\}$, are generators of a group if any group element can be formed from linear combinations of them. There are at least $N^2 - 1$ such generators, where N is the dimension of the group. The generators are not necessarily a basis, as they do not need to be linearly independent of each other. For groups such as the SM gauge groups[2], any element $Q(x)$, can be represented in the form,

$$Q(x) = \exp\left(i \sum_{j=1}^{N^2-1} \alpha_j(x) T_j \right), \tag{2.2}$$

where $\alpha_j(x)$ are complex coefficients. The generators satisfy the commutation relation,

$$[T_i, T_j] = i f_{ijk} T_k, \tag{2.3}$$

where f_{ijk} are known as "structure constants". For an abelian group, where all elements commute, the structure constants are given by, $f_{ijk} = 0$.

The generators of an SM gauge group are understood to be the fundamental representations of the bosons, also known as "gauge bosons". The minimum

[1] More specifically, Lie groups.

[2] Lie groups.

number of generators of a gauge group determines the number of distinct boson fields. The structure constants of the group determine how the boson fields interact with each other. The form of the group determines how the fermion fields transform and thus how they interact with the boson fields. Or from the historical perspective, the non-trivial transformations of the fermion fields under which the Lagrangian is invariant, necessarily produce extra fields with well defined couplings to the fermions and with each other. In the next section, the appearance of these gauge groups in the SM is outlined.

2.1.3 Lagrangian formulation and bosons

Noether's theorem says that the conserved quantities of a physical system result from its symmetries, and thus the fermion equations of motion are examined for symmetries in this section. The most powerful way to associate equations of motions, symmetries and conserved quantities is through the Lagrangian formulation. Similar to the classical mechanics case, the equations of motion of a system can be determined from the Euler-Lagrange equations,

$$\frac{\partial}{\partial x_\mu}\left(\frac{\partial \mathcal{L}}{\partial(\partial\psi/\partial x_\mu)}\right) - \frac{\partial \mathcal{L}}{\partial \psi} = 0, \tag{2.4}$$

where $\psi = \psi(x_\mu, t)$ is a field and given the Lagrangian density $\mathcal{L}$, which is referred to as the Lagrangian in what follows.

Transformations of the field that do not change the form of the Lagrangian are of interest for finding conserved quantities. Consider the Lagrangian,

$$\mathcal{L} = i\bar{\psi}\gamma_\mu\partial^\mu\psi - m\bar{\psi}\psi, \tag{2.5}$$

which produces the Dirac equation of Eq. (2.1), when input into the Euler-Lagrange equations. The form of this Lagrangian is chosen precisely to produce the Dirac equation, and is otherwise arbitrary. Yet, given the significance of the Dirac equation, much can be learned from the form of this Lagrangian. The two terms, quadratic in the field, represent the propagation of a massive particle (i.e. a fermion) through empty space via a kinetic energy. To produce a symmetry consider the transformation,

$$\psi(x_\mu) \to e^{i\alpha}\psi(x_\mu), \tag{2.6}$$

where α is some arbitrary constant, and the transformation forms the $U(1)$ gauge group due to Eq. (2.2). The Lagrangian is invariant under this transformation

Table 2.3: Properties of the bosons, taken from [9], where Q is the electric charge, I_z, the z component of weak isospin, and r, g and b are red, green and blue colour charges.

Name	Symbol	Mass (MeV/c^2)	Gauge Group	Q	I_z	Colour charge
Photon	γ	0	$U(1)$	0	0	0
Neutral Weak	Z^0	91,187.6 $\pm$ 2.1	$SU(2) \times U(1)$	0	0	0
Charged Weak	$W^{\pm}$	80,385 $\pm$ 15	$SU(2) \times U(1)$	± 1	± 1	0
Gluon	g	0	$SU(3)$	0	0	r,g,b

and thus has a conserved quantity, which in the case of electromagnetism is the current density. It follows that the electric charge must be a conserved quantity.

Now consider the most general case where $\alpha := \alpha(x_\mu)$, where the constant is now an arbitrary function of space-time position. The Lagrangian of Eq. (2.5) is not invariant under this transformation due to the additional $\partial_\mu \alpha(x_\mu)$ term. By introducing a vector field, a new Lagrangian is constructed that is indeed invariant,

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu Q \psi A_\mu, \tag{2.7}$$

where Q is the electric charge of the fermion, e is a coupling constant and the new field A_μ transforms as,

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha(x_\mu). \tag{2.8}$$

The new term in the Lagrangian shows the interaction between the field ψ and the new vector field A_μ. The value of the coefficient eQ, gives the strength of the interaction coupling while the matrix form of γ^μ encodes the spin of the fermion before and after the interaction, and determines which interactions are possible.

Of course the form of the previous equations are chosen to reproduce the form of the electromagnetic interaction, with the vector field A_μ understood to be the photon field, the force carrying boson of electromagnetism. Table 2.3 shows the properties of the SM bosons for the three fundamental forces. To complete this treatment, a kinetic energy propagation term for the photon is needed. The final Lagrangian for quantum electrodynamics (QED) is,

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu Q \psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.9}$$

where the gauge invariant field strength tensor is given by,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.10}$$

A boson kinetic term of this form necessarily arises in an abelian gauge group, as with $U(1)$ for electromagnetism [14].

The "local" gauge transformation, $\alpha := \alpha(x_\mu)$, is not a symmetry of the system if a mass term ($\propto A_\mu A^\mu$) is added to the photon propagator. A finite mass indicates a finite interaction range, which can't hope to correct for an infinite number of local transformations over all space. Thus the photon is a massless boson.

For the strong nuclear force there are extra degrees of freedom introduced by the three distinct values of colour charge, as opposed to one for conventional electric charge. This suggests the form of the strong interaction is more complex than that of the EM force. It is found that the $SU(3)$ group describes the observed features of the strong force through the quark field transformation,

$$q(x_\mu) \to e^{i\alpha_a T_a} q(x_\mu), \tag{2.11}$$

where T_a are the generators of $SU(3)$ and α_a are coefficients. The generators for the strong force are conventionally chosen to be $\lambda_a/2$, where the λ_a are the Gell-Mann matrices. In analogy to the previous strategy, a new field G^a_μ, must be added to the Lagrangian. The final Lagrangian for quantum chromodynamics (QCD), also including a boson field kinetic term, is given by,

$$\mathcal{L} = \bar{q}(i\gamma^\mu \partial_\mu - m)q + g_s(\bar{q}\gamma^\mu T_a q)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a. \tag{2.12}$$

In this case the transformation of the field in Eq. (2.8) is not sufficient to preserve gauge invariance due to the non-commuting generators adding an extra term. Consequently an additional term is needed in the transformations that includes the structure constants,

$$G^a_\mu \to G^a_\mu + \frac{1}{g_s}\partial_\mu \alpha_a - f_{abc}\alpha_b G^c_\mu. \tag{2.13}$$

The field strength tensor also includes an extra term,

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f_{abc} G^b_\mu G^c_\nu. \tag{2.14}$$

Thus the QCD Lagrangian is formed from quark fields that interact via $N^2 - 1 = 8$ vector bosons called "gluons". The gluons couple to particles that carry colour charge. The extra terms arising from the non-abelian nature of the $SU(3)$ colour gauge group, such as $(\partial_\mu G^a_\nu)G^\mu_b G^\nu_c$ result in gluon self interactions. Due to these terms, triple and quadruple gluon vertices are possible in Feynman diagrams, which suggests that gluons themselves carry colour charge.

2.1.4 Higgs mechanism

The motion of a massive spin-0 particle is described by the Klein-Gordon equation [16],

$$\partial_\nu \partial^\nu \phi + \mu^2 \phi = 0, \tag{2.15}$$

and is reproduced by the Lagrangian,

$$\mathcal{L} = \frac{1}{2}(\partial_\nu \phi)(\partial^\nu \phi) - \frac{1}{2} m^2 \phi^2. \tag{2.16}$$

Consider the modified Klein-Gordon Lagrangian to which a boson field, A_ν, has been added in analogy to previous sections, along with a quartic function ansatz,

$$\mathcal{L} = (\partial_\nu - ieA_\nu)\phi^*(\partial_\nu + ieA_\nu)\phi - \frac{1}{4}F_{\nu\xi}F^{\nu\xi} - \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2. \tag{2.17}$$

The last two terms are a potential $V(\phi)$ and ϕ is a complex scalar field with real components of the form,

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2). \tag{2.18}$$

If $\mu^2 < 0$ then the potential $V(\phi)$ is the "wine bottle" potential, which has an infinite number of minima in a circle around the origin, and has a local maximum at the origin. If the system is initially at the local maximum, it will spontaneously evolve into one of an infinite number of possible states, depending on chaotic initial conditions. A gauge transformation of the Lagrangian then moves the system from one degenerate minimum to another. By choosing one of the arbitrary degenerate minima, thereby favouring one direction over the other infinite directions, the gauge symmetry of the system is "spontaneously broken". The values of the fields in the minima, or ground state, are chosen to be, $\phi_1 = v$ and $\phi_2 = 0$, where the loss of generality is explicit. Of interest are excitations of the field about this special minimum,

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \zeta(x) + i\eta(x)). \tag{2.19}$$

The Lagrangian is then [14],

$$\mathcal{L} = \frac{1}{2}(\partial_\nu \eta)^2 + \frac{1}{2}(\partial_\nu \zeta)^2 + \frac{1}{2}e^2 A^\nu A_\nu v^2 + evA^\nu \partial_\nu \eta + \mu^2 \zeta^2 + \ldots \tag{2.20}$$

The mass term, $\mu^2\zeta^2$, shows that a field ζ, with mass $m_\zeta = \sqrt{-2\mu^2}$ has been produced. Interestingly a boson mass term, $\frac{1}{2}e^2 v^2 A^\nu A_\nu$ is also produced. Thus the mechanism of spontaneous symmetry breaking gives mass to the gauge boson in this way. The field ζ is understood to be that representing the "Higgs boson", named after one of the proponents of the mechanism [17,18]. No mass term exists for the η field, known as a Goldstone field, after Goldstone's theorem. It can be removed by a choice of gauge and thus doesn't contribute to the physical theory. The Goldstone field arises due to excitations of the field in the direction of the other infinite minima.

An explicit mass term for fields split into their left- and right-handed parts, breaks the invariance of the Lagrangian under $SU(2)$ transformations, which are necessary for the weak force to be included in the SM. Instead the fermion mass term from previous equations is removed and replaced with a Yukawa term [14],

$$\mathcal{L}_{\text{Yukawa}} = -Y_\psi \left[\bar{\psi}_L \phi \psi_R + h.c\right], \tag{2.21}$$

where Y_ψ is the "Yukawa coupling" for fermion ψ, L and R represent the definite handedness parts, and $h.c$ is the Hermitian conjugate of the first term. The fermions acquire their mass through this term when the symmetry is broken as before. It also acquires interaction terms with the Higgs field, ζ.

2.1.5 Electroweak bosons

Consider the $SU(2) \times U(1)$ composite group. Different transformations for the left- and right-handed fields can be freely chosen,

$$\begin{aligned} \psi_L &\to e^{i\alpha_a T_a + i\beta Y} \psi_L, \\ \psi_R &\to e^{i\beta Y} \psi_R. \end{aligned} \tag{2.22}$$

These are chosen to reproduce the behaviour of the weak nuclear force that does not couple to right-handed particles. Here, T_a are the three generators of $SU(2)$, which by convention are the three Pauli matrices. In the transformations, Y is the "weak hypercharge", the conserved quantity of the $U(1)$ part of the electroweak force. The conserved quantity of the $SU(2)$ part of the electroweak force is the "weak isospin", the quantum number for which is defined along the z component axis, I_z. The weak hypercharge is given by,

$$Y = 2(Q - I_z). \tag{2.23}$$

In analogy to previous sections, four new gauge boson fields must be added for the Lagrangian to be invariant under the above transformation, three for $SU(2)$ and one for $U(1)$. It is seen experimentally that the weak force carrier bosons are massive, while the photon of $U(1)$ is massless as discussed previously. The Higgs spontaneous symmetry breaking direction can be chosen arbitrarily, and the four new bosons can be combined in such a way as to form three massive bosons and one massless boson. The "charged weak current" bosons, $W^{\pm}$ and the "neutral weak current" Z^0 boson, carry the weak nuclear force, while the fourth electroweak boson is the photon that carries only the $U(1)$ EM force. Due to the non-abelian nature of the $SU(2)$ group, the electroweak bosons can couple to each other, and form three and four point Feynman vertices.

2.2 $\hat{C}$, $\hat{P}$ and $\hat{T}$ operators

This section details three relevant operators, the symmetries of which are investigated in this thesis. These are the parity operator ($\hat{P}$), the charge-conjugation operator ($\hat{C}$) and the time-reversal operator ($\hat{T}$).

The parity operator acts on a state and returns a state with reflected spacial coordinates,

$$\hat{P}\psi(\vec{r},t) = P\psi(-\vec{r},t). \tag{2.24}$$

A second operation gives the possible eigenvalues of the operator,

$$\hat{P}^2\psi(\vec{r},t) = P\hat{P}\psi(-\vec{r},t) = P^2\psi(\vec{r},t), \tag{2.25}$$

which gives $P^2 = 1$. P is an observable and thus $\hat{P}$ is a Hermitian operator with real eigenvalues, so that $P = \pm 1$. The full SM Lagrangian is not invariant under the parity transformation as weak nuclear interactions never conserve parity.

The charge conjugation operator changes a particle into its anti-particle. In this transformation a particle's position and momentum are unchanged. Consider a particle p with anti-particle $\bar{p}$ and a particle q that is its own anti-particle. The $\hat{C}$ operator acts on these states as,

$$\hat{C}|p\rangle = |\bar{p}\rangle, \qquad \hat{C}|q\rangle = C_q|q\rangle. \tag{2.26}$$

A second operation gives $C = \pm 1$. There is no factor for the operation on p as the final state is not the same as the initial state and thus $|p\rangle$ is not an eigenvector of $\hat{C}$. Similarly to the parity operator, $\hat{C}$ is never conserved by the weak nuclear interaction.

Time reversal is the reflection of a state's time coordinate. In classical mechanics the $\hat{T}$ operator changes a particle's momentum direction, where $\bar{t} = -t$,

$$\vec{p} = m\frac{\partial}{\partial t}\vec{x}(t) \mapsto m\frac{\partial \bar{t}}{\partial t}\frac{\partial}{\partial \bar{t}}\vec{x}(-t) = -m\frac{\partial}{\partial \bar{t}}\vec{x}(\bar{t}) = -\vec{\bar{p}}.$$

By invoking the correspondence principle, this must also hold in quantum mechanics. Consider a plane wave state with momentum, $\vec{p}$ and position $\vec{r}$,

$$\Psi_{\vec{p}}(\vec{r},t) = e^{\frac{i}{\hbar}(\vec{p}\cdot\vec{r}-Et)}.$$

Time reversal gives,

$$\Psi_{\vec{p}}(\vec{r},t) \mapsto \Psi_{-\vec{p}}(\vec{r},-t) = e^{\frac{i}{\hbar}((-\vec{p})\cdot\vec{r}+Et)} = \Psi^{*}_{\vec{p}}(\vec{r},t), \tag{2.27}$$

and thus time reversal can be represented by complex conjugation. Note that time reversal is not Hermitian nor is it a linear map. Thus $\hat{T}$ has no associated observable and is not strictly an operator.

The joint operation of $\hat{C}$ and $\hat{P}$ is conserved in the strong and electromagnetic interactions. It is known to be violated in the weak interaction, specifically in the K and B meson systems, this is known as *CP* violation (CPV) and is discussed in the next sections. The joint operation of $\hat{C}\hat{P}\hat{T}$ is thought to be exact in nature. This symmetry is linked to Lorentz transformations and thus a violation of this symmetry is a violation of Lorentz invariance. As a result, the operation of $\hat{C}\hat{P}$ is equivalent but opposite to that of $\hat{T}$. As the operation of $\hat{T}$ is equivalent to complex conjugation, a violation of the $\hat{T}$ symmetry, and thus the $\hat{C}\hat{P}$ symmetry, requires complex terms in the SM Lagrangian.

2.3 Mixing

When more than one fermion generation is considered, the Yukawa couplings of Eq. (2.21) become matrices encoding the couplings for each generation. These matrices relate the flavour states of the leptons and quarks to their mass states. This means that the couplings between, say the u and d quarks, in the electroweak interaction are not necessarily the same as their couplings in the Yukawa terms. In other words, the interaction states of the fermions and their definite mass states may not be the same representations.

This was first parameterised by Cabibbo in 1963 [19], using only the first two generations of quarks and a unitary transformation between the two bases,

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}, \tag{2.28}$$

where d' and s' are the definite mass states and θ_C is the Cabibbo angle. This was proposed when only three quarks had been discovered, however a fourth quark is needed as the up-type partner of the s quark. Thus the work of Cabibbo was one of a number of formalisms that led to the prediction of the c quark. The square of the transformation matrix elements are proportional to the coupling strengths to the electroweak interaction. Note that there is no reason why the parameterisation cannot be in terms of the u or c quarks instead.

It is found that the transformation is not a diagonal matrix such that quarks can interact between generations. Thus unlike strong interactions, these weak interactions can occur between quarks from different generations. This has interesting consequences for meson lifetimes. Even when a weak nuclear decay is allowed it will have a relatively long lifetime, which is a consequence of the large mass of the weak bosons. An allowed strong decay is far more likely than a weak decay, which will only dominate when the strong decay is forbidden by various conservation laws. As a result, the lowest energy meson from each of the quark families has a relatively long lifetime, as there are no available strong decays. In this case, weak decays that violate quark flavour are allowed, but occur less frequently. This also results in a characteristic cascade of decays down the quark flavour families, such as $B \to D \to K$, and so on.

Experimental measurements of the Cabibbo angle find that $\theta_C = (13.01 \pm 0.03)^\circ$ [9]. Thus the diagonal couplings are favoured with respect to the off-diagonal couplings between the two generations. Interactions involving the diagonal elements are "Cabibbo favoured" (CF) while those involving the off-diagonal terms are "Cabibbo suppressed" (CS).

The violation of *CP* symmetry was discovered in 1964 [20], and is discussed in the next section. In 1973, Kobayashi and Maskawa looked to the Cabibbo matrix to include *CP* violation in the SM [21]. As discussed in Section 2.2, *CP* violation can only occur in complex terms of the SM Lagrangian. A unitary, and thus orthogonal, matrix contains $\frac{1}{2}(n-1)(n-2)$ complex phases, where n is the number of fermion generations (dimensions) [22]. For $n = 2$, the Cabibbo matrix contains no complex phases. Kobayashi and Maskawa predicted an additional quark generation such that one additional dimension is added to the transformation matrix, which now has a complex phase. This new transformation matrix is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix,

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.29}$$

Again, the off-diagonal elements are suppressed as they are relatively small compared to the diagonal elements. This can be seen more clearly in the Wolfenstein parameterisation [22],

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} 1-\beta^2/2 & \beta & A\beta^3(\rho - i\eta) \\ -\beta & 1-\beta^2/2 & A\beta^2 \\ A\beta^3(1-\rho-i\eta) & -A\beta^2 & 1 \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.30}$$

where terms of order β^4 are neglected. The values of the parameters are found to be $\beta = \sin\theta_C \approx 0.23$, $A \approx 0.81$ and $\rho - i\eta = 0.13 - 0.37i$ [9].

For neutral mesons this transformation matrix results in the phenomenon of mixing. This is the suppressed process by which a meson can "mix" into its anti-partner as shown in Fig. 2.1 for D^0 mixing. Mixing processes are known to occur in the K^0, B^0, B_s^0 and D^0 meson systems. The exact same Lagrangian terms for the leptons also results in neutrino flavour mixing, if the neutrinos are prescribed a mass. The "box" diagram of Fig. 2.1 is a highly suppressed process, as it has an amplitude proportional to β^4 at largest, depending on the intermediate quark involved. The diagram is also suppressed by the GIM mechanism [23]. The dominant process of D^0 meson mixing is the long-range contribution of Fig. 2.2, in which the initial D^0 decays into a state accessible by both D^0 flavours, such as $\pi^+\pi^-$ or K^+K^-. The $\overline{D}^0$ is then formed by recombination.

To investigate mixing in terms of the probability of a state interacting in its matter or anti-matter state, the mass eigenstates are first written as a linear combination of the definite flavour states,

$$|D^0{}_{1,2}\rangle = p|D^0\rangle \pm q|\overline{D}^0\rangle, \tag{2.31}$$

where p and q are complex coefficients with $|p|^2 + |q|^2 = 1$. The time evolution of these states is determined by the Hamiltonian for the weak interaction and the

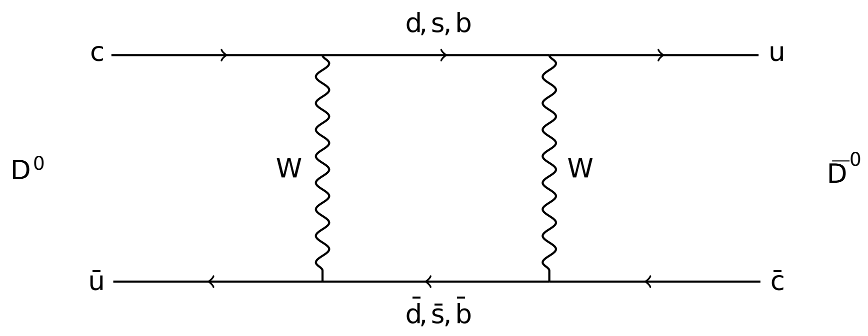


Figure 2.1: Feynman box diagram of D^0-$\overline{D}^0$ mixing [24].

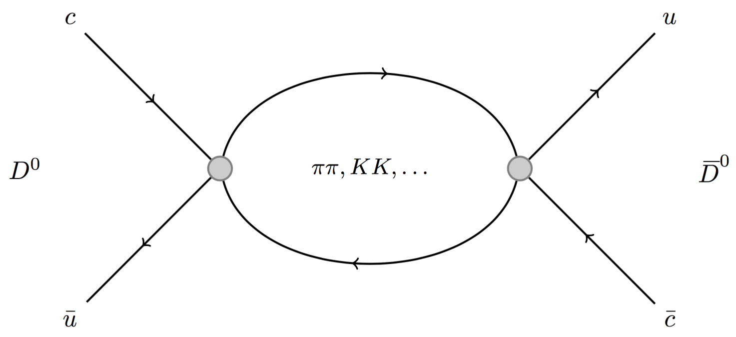


Figure 2.2: Long range contribution to D^0-$\overline{D}^0$ mixing [24].

time-dependent Schrödinger equation [25, 26],

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = \boldsymbol{H}|\Psi(t)\rangle = \left(\boldsymbol{M} - \frac{i}{2}\boldsymbol{\Gamma}\right)|\Psi(t)\rangle, \tag{2.32}$$

where the non-Hermitian complex Hamiltonian matrix has been decomposed into Hermitian mass and decay width matrices and,

$$|\Psi(t)\rangle = \begin{pmatrix} D^0(t) \\ \overline{D}^0(t) \end{pmatrix}. \tag{2.33}$$

The solution to the eigenvalue problem is given by [25],

$$\begin{aligned} |\Psi(t; D^0)\rangle = \frac{1}{2}\Big[&(e^{-im_2t-\frac{1}{2}\Gamma_2 t} + e^{-im_1t-\frac{1}{2}\Gamma_1 t})|D^0\rangle \\ &+ \frac{q}{p}(e^{-im_2t-\frac{1}{2}\Gamma_2 t} - e^{-im_1t-\frac{1}{2}\Gamma_1 t})|\overline{D}^0\rangle\Big], \end{aligned} \tag{2.34}$$

for an initial state $|D^0\rangle$, where m_i and Γ_i are the eigenvalues of $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ for the two mass states. A similar expression exists for the state that is purely $|\overline{D}^0\rangle$ at $t = 0$. The probability of an initial $|D^0\rangle$ state interacting as a $|\overline{D}^0\rangle$ state at some time t is given by,

$$|\langle\overline{D}^0|\boldsymbol{H}|\Psi(t; D^0)\rangle|^2 = \frac{1}{2}\left|\frac{q}{p}\right|^2 e^{-\tau}\left(\cosh(\tau y) - \cos(\tau x)\right), \tag{2.35}$$

and for a state that is initially $|\overline{D}^0\rangle$ at t=0, the probability of interacting as $|D^0\rangle$ is,

$$|\langle D^0|\boldsymbol{H}|\Psi(t; \overline{D}^0)\rangle|^2 = \frac{1}{2}\left|\frac{p}{q}\right|^2 e^{-\tau}\left(\cosh(\tau y) - \cos(\tau x)\right), \tag{2.36}$$

where the mixing parameters are defined as,

$$x := \frac{m_2 - m_1}{\Gamma}, \qquad y := \frac{\Gamma_2 - \Gamma_1}{2\Gamma}, \tag{2.37}$$

with average decay width, $\Gamma := (\Gamma_1 + \Gamma_2)/2$ and time, $\tau := \Gamma t$.

2.4 Violation of the *CP* symmetry

Comparing Eqs. (2.35) and (2.36) shows that the probabilities of pure initial states interacting as matter or anti-matter are different if,

$$\left|\frac{q}{p}\right| \neq 1, \tag{2.38}$$

which would indicate a clear preference in nature for interactions in either the matter state or the anti-matter partner state, and would be unambiguous evidence of *CP* violation. This is one type of *CP* violation that is due to a difference in the mixing rates of D^0 and $\overline{D}^0$ and is known as *CP* violation in mixing. To investigate the other two forms of *CP* violation, the full decay rate for a D^0 decaying to a *CP* even final state f is [25],

$$\begin{aligned}\Gamma(D^0(t) \to f) &:= |\langle f|\boldsymbol{H}|\Psi(t;D^0)\rangle|^2, \\ &= \frac{1}{2}e^{-\Gamma t}|A_f|^2\Big[(1+|\lambda_f|^2)\cosh(y\tau) + (1+|\lambda_f|^2)\cos(x\tau) \\ &+ 2\mathrm{Re}(\lambda_f)\sinh(y\tau) - 2\mathrm{Im}(\lambda_f)\sin(x\tau)\Big],\end{aligned} \tag{2.39}$$

where $A_f := \langle f|\boldsymbol{H}|D^0\rangle$ and $\bar{A}_f := \langle f|\boldsymbol{H}|\overline{D}^0\rangle$ are the transition amplitudes and,

$$\lambda_f := \frac{q}{p}\frac{\bar{A}_f}{A_f} = -\left|\frac{q}{p}\right|\left|\frac{\bar{A}_f}{A_f}\right|e^{i\phi}, \tag{2.40}$$

where the complex part of the first equality is absorbed into the "weak phase", ϕ. A similar decay rate, $\Gamma(\overline{D}^0(t) \to f)$, for the initial $|\overline{D}^0\rangle$ state shows that the rates are different if $|\lambda_f| \neq 1$. This can occur in three distinct ways, where the first was discussed above. If the decay amplitudes are different, then one of the decays to the final state f is favoured over the partner's decay to f. This is known as "direct" *CP* violation or *CP* violation in decay. Finally, if the weak phase is not zero then there is interference between the first two types of *CP* violation. The *CP* violation in mixing and interference are collectively called "indirect" *CP* violation.

Violation of the *CP* symmetry has been observed in the K, B, B_s^0 and B^+ systems [20, 27, 28]. All current measurements of the D system show no evidence of *CP* violation. The first discovery of CPV was made by observing *CP* violating decays of the K_S and K_L mesons, which are the mass eigenstates [20]. These two mesons have similar masses but very different lifetimes, resulting in a large $\Delta\Gamma = \Gamma_2 - \Gamma_1$, which gives a large probability for an initial state of one of these particles to interact as the other. It was found that a beam of pure K_L could decay to two pions, which violated *CP* symmetry. This suggests that the two meson states are linear combinations of the true *CP* and flavour eigenstates, K^0, $\overline{K}^0$, with very different linear coefficients. Thus in the K^0 system there is indirect CPV, $|q/p| \neq 1$. There is also direct CPV in this system but it is found to be much smaller than the indirect effect, due to the small order for which complex terms appear in kaon decay quark couplings as in Eq. (2.30). It is the large width difference between the mass eigenstates that makes kaon CPV large enough to be observed.

For B meson decays the direct CPV effect is largest as $\Delta\Gamma$ is much smaller than in the kaon system, but the B meson decays have access to the $|V_{ub}|^2$ coupling in Eq. (2.30). This coupling is of order β^6, whereas the other accessible couplings of the complex coefficient are of order smaller than β^8.

The apparent lack of CPV in the D systems can be explained as it neither has a large amount of mixing, nor access to the larger order complex terms in the CKM quark couplings. The $|V_{us}|^2$ and $|V_{cd}|^2$ couplings involved in D^0 decays only contain the potentially *CP* violating complex part at an order below β^8, and the lifetime difference and mixing parameters are very small [9].

2.5 *CP* violation observables

Of greatest interest to this thesis is the indirect violation of *CP*. Observables that parameterise these effects are constructed in this section.

The mixing parameters x and y are not yet well measured due to the small difference between the masses and widths of the D^0 mass states. In contrast the K system has a very large width difference, which allows y to be more easily measured. The D^0 mixing parameters are known to be of order 0.01 [9], as shown in Fig. 2.3. As a result, the trigonometric and hyperbolic terms in Eq. (2.39) can be Taylor expanded to first order, giving:

$$\Gamma(D^0(t) \to f) = e^{-\Gamma t}|A_f|^2 \left[1 - \left|\frac{q}{p}\right|\left|\frac{\bar{A}_f}{A_f}\right|(y\cos\phi - x\sin\phi)\Gamma t\right], \tag{2.41}$$

$$\Gamma(\overline{D}^0(t) \to f) = e^{-\Gamma t}|\bar{A}_f|^2 \left[1 - \left|\frac{p}{q}\right|\left|\frac{A_f}{\bar{A}_f}\right|(y\cos\phi + x\sin\phi)\Gamma t\right], \tag{2.42}$$

with similar expressions for a *CP* odd final state $\bar{f}$. The above equation is modelled as a single exponential in the analysis to follow, thus it can be written in terms of an effective decay width. For the $|D^0\rangle$ initial state this is,

$$\Gamma(D^0(t) \to f) := |A_f|^2 e^{-\hat{\Gamma} t}, \tag{2.43}$$

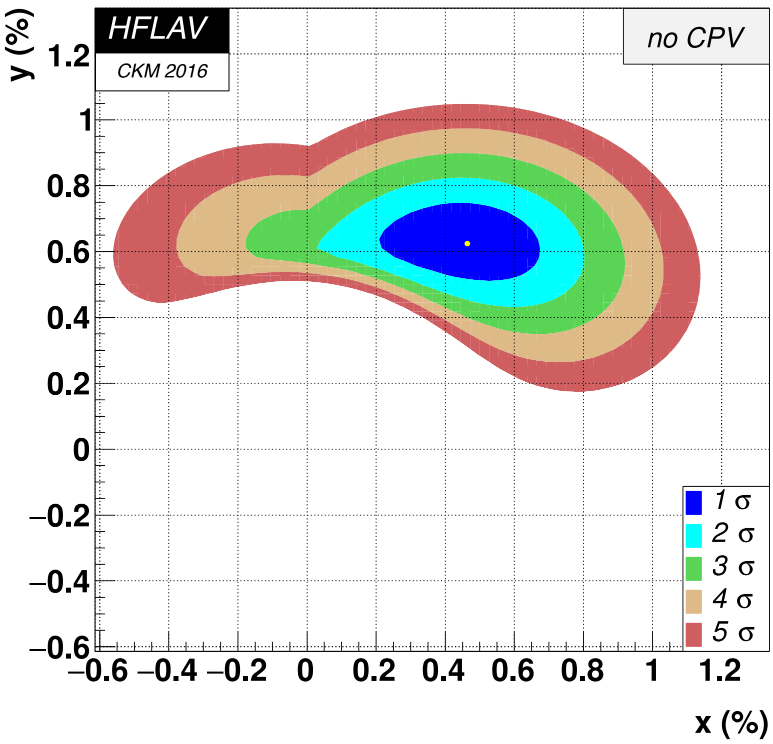


Figure 2.3: HFLAV average of the D^0 mixing parameters x and y from the CKM 2016 conference and assuming no CP violation [6].

where the effective decay time is related by $\hat{\tau} = 1/\hat{\Gamma}$. The square brackets of Eq. (2.41) are a first order expansion of an exponential, for small x and y, and can be written as,

$$\Gamma(D^0(t) \to f) = e^{-\Gamma t}|A_f|^2 \exp\left[-\left|\frac{q}{p}\right|\left|\frac{\bar{A}_f}{A_f}\right|(y\cos\phi - x\sin\phi)\Gamma t\right], \tag{2.44}$$

so that the effective decay width is,

$$\hat{\Gamma}(D^0 \to f) = \Gamma\left[1 + \left|\frac{q}{p}\right|\left|\frac{\bar{A}_f}{A_f}\right|(y\cos\phi - x\sin\phi)\right]. \tag{2.45}$$

The expected small size of CP violation in the D^0 system allows the assumption that $|q/p|$ is responsible for a small CPV in mixing effect and similarly $|\bar{A}_f/A_f|$ has a small component due to direct CPV,

$$\left|\frac{p}{q}\right|^{\pm 2} \approx (1 \pm A_m), \qquad \left|\frac{\bar{A}_f}{A_f}\right|^{\pm 2} \approx (1 \pm A_d). \tag{2.46}$$

The observables of interest are constructed as a sum and a difference of the effective widths,

$$y_{CP} = \frac{\hat{\Gamma} + \hat{\bar{\Gamma}}}{2\Gamma} - 1, \qquad A_\Gamma = \frac{\hat{\Gamma} - \hat{\bar{\Gamma}}}{\hat{\Gamma} + \hat{\bar{\Gamma}}}. \tag{2.47}$$

In terms of the effective lifetimes, which are measured in this thesis, A_Γ is given by,

$$A_\Gamma = \frac{\bar{\tau} - \tau}{\bar{\tau} - \tau}, \tag{2.48}$$

where the hat symbol has been omitted for clarity. The difference and sum of the effective widths are given by,

$$\hat{\Gamma} - \hat{\bar{\Gamma}} \approx 2\Gamma \left\{ \frac{1}{2}(A_m + A_d) y \cos\phi - \left(1 - \frac{1}{8}(A_m^2 - 2A_m A_d)\right) x \sin\phi \right\}, \tag{2.49}$$
$$\hat{\Gamma} + \hat{\bar{\Gamma}} \approx 2\Gamma \left\{ -\frac{1}{2}(A_m + A_d) x \sin\phi + \left(1 - \frac{1}{8}(A_m^2 - 2A_m A_d)\right) y \cos\phi + 1 \right\}.$$

Then the observables are,

$$y_{CP} \approx \left\{ -\frac{1}{2}(A_m + A_d) x \sin\phi + \left(1 - \frac{1}{8}(A_m^2 - 2A_m A_d)\right) y \cos\phi \right\}, \tag{2.50}$$
$$A_\Gamma \approx \frac{1}{y_{CP} + 1} \left\{ \frac{1}{2}(A_m + A_d) y \cos\phi - \left(1 - \frac{1}{8}(A_m^2 - 2A_m A_d)\right) x \sin\phi \right\}.$$

Some of these parameters can be neglected due to their relative sizes from experimental measurements and constraints [6, 29]. The $\sin(\phi)$ term is of order 10^{-1}, while the A_m and A_d terms have maximum orders of 10^{-1} and 10^{-2}, respectively. Only terms of order 10^{-2} are kept such that the equations reduce to,

$$y_{CP} \approx \left\{ -\frac{1}{2} A_m x \sin\phi + \left(1 - \frac{1}{8} A_m^2\right) y \cos\phi \right\}, \tag{2.51}$$
$$A_\Gamma \approx \frac{1}{y_{CP} + 1} \left\{ \frac{1}{2}(A_m + A_d) y \cos\phi - x \sin\phi \right\}. \tag{2.52}$$

The observable y_{CP} is predominantly a measure of the mixing parameter y, in the absence of CPV. Any difference between the mixing parameter y and y_{CP},

Table 2.4: Recent results for A_Γ extracted from $D^0 \to h^+h^-$ decays. The errors quoted are statistical and systematic uncertainties, respectively.

Collaboration	$A_\Gamma(K^+K^-)(10^{-3})$	$A_\Gamma(\pi^+\pi^-)(10^{-3})$	Combined(10^{-3})
Belle [30]	-	-	$-0.3 \pm 2.0 \pm 0.8$
BaBar [31]	-	-	$0.9 \pm 2.6 \pm 0.6$
LHCb (2010) [32]	$-5.9 \pm 5.9 \pm 2.1$	-	-
LHCb (2011) [33]	$-0.35 \pm 0.62 \pm 0.12$	$0.33 \pm 1.06 \pm 0.14$	-
CDF(2014) [34]	$-1.9 \pm 1.5 \pm 0.4$	$-0.1 \pm 1.8 \pm 0.3$	-1.2 ± 1.2
LHCb (μ tag) [35]	$-1.34 \pm 0.77^{+0.26}_{-0.34}$	$-0.92 \pm 1.45^{+0.25}_{-0.33}$	-1.25 ± 0.73

indicates CPV. The A_Γ observable is a measure of indirect CPV and must be consistent with zero to preserve *CP* symmetry. The observable of interest to this thesis is thus A_Γ, and its measurement is described in detail in Part II. Previous measurements of A_Γ are shown in Table 2.4, where the first errors are statistical and the second are systematic. All current measurements of A_Γ are consistent with zero and with *CP* conservation in the D system.

In terms of the final states of interest in this thesis, the average width Γ is the decay width of the Cabibbo favoured (CF) decay $D^0 \to K^-\pi^+$. No *CP* violating terms enter into its decay amplitudes, and thus A_Γ measured with this decay is expected to be zero. The effective widths, such as Eq. (2.45), are determined in the *CP* even, singly Cabibbo suppressed (SCS) final states $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ by measuring their effective lifetimes, $\hat{\tau}$.

Chapter 3

Experiment apparatus

The data analysed in this thesis were collected using the LHC beauty (LHCb) detector, which is one of seven experiments taking data from collisions at the Large Hadron Collider (LHC), located on the French-Swiss border near Geneva. The detector is operated by the LHCb collaboration consisting of about 1,400 members, and the collider is operated by the European Organisation for Nuclear Research (CERN). The accelerator chain of the LHC and its design are described in Section 3.1. Detailed descriptions of the LHCb sub-detectors and their performance are given in Section 3.2.

3.1 Large Hadron Collider

The LHC [36, 37] is a circular proton-proton (pp) collider with a circumference of 26.7 km and a design centre-of-mass energy ($\sqrt{s}$) of 14 TeV. The centre-of-mass-energy is a measure of the energy of the colliding protons in the centre-of-mass frame. The LHC has a design luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and a design collision rate of 40 MHz. It began operations in 2008 and started taking data in late 2010 and throughout 2011 at an energy of $\sqrt{s} = 7\,\mathrm{TeV}$ and a collision rate of 20 MHz. In 2012 the machine took data at a centre-of-mass energy of 8 TeV before being shut down in early 2013 for two years of repairs and upgrades to allow for higher energies. Since 2015 the machine has been running at the design collision rate with a bunch spacing time of 25 ns and a centre-of-mass energy of 13 TeV. The first running period of the machine from 2010 to 2013 is known as Run 1 and the second running period from 2015 to 2018 is Run 2.

Six other experiments directly use collisions from the LHC. The high luminosity general purpose detectors ATLAS [38] and CMS [39], use the maxi-

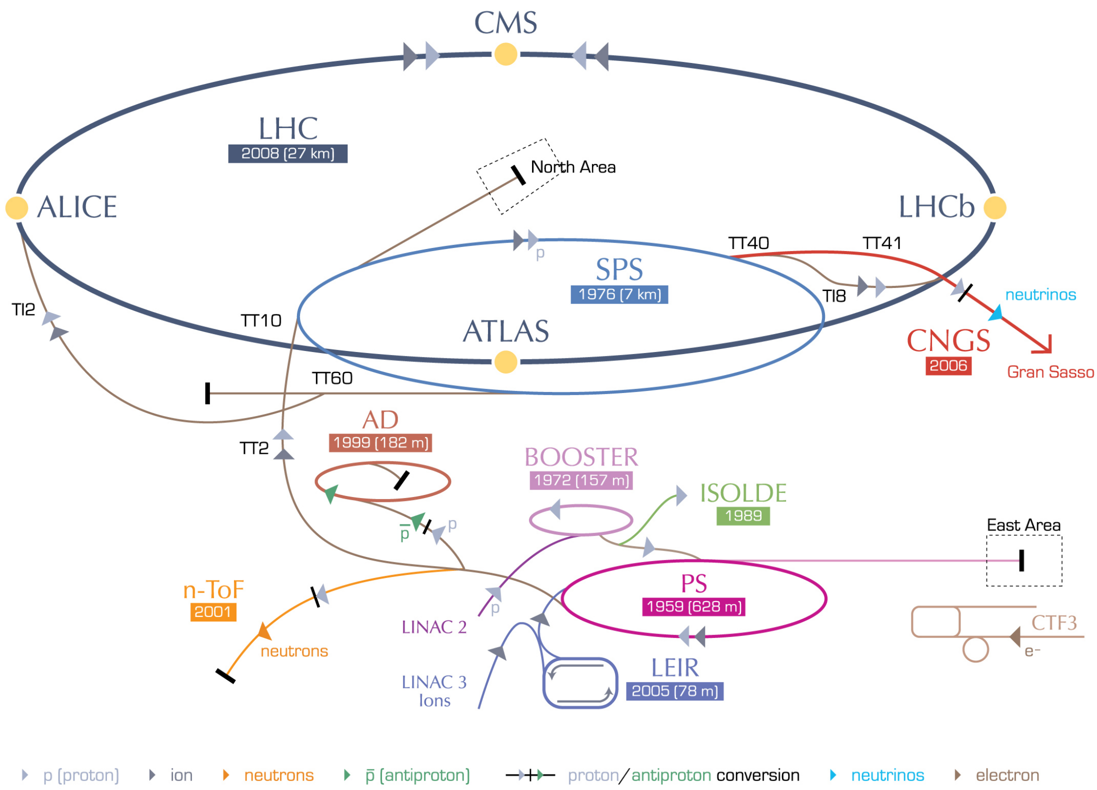


Figure 3.1: Schematic of the CERN accelerator complex, showing the LHC accelerator chain [44].

mum luminosity provided by the LHC. The dedicated heavy ion experiment, ALICE [40] studies quark-gluon plasma. There are also three smaller experiments, MoEDAL [41], which searches for magnetic monopoles, LHCf [42], which studies particles produced at small angles with respect to the beamline ("forward region") and TOTEM [43], which measures cross-sections and scattering processes in the forward region.

The chain of accelerators supplying protons to the LHC is shown in Fig. 3.1. The initial protons come from a hydrogen gas source that is ionised to strip away the electrons. The first accelerator in the chain is the linear accelerator, LINAC2, which accelerates the protons to an energy of 50 MeV. The next stage in the chain is the proton synchrotron booster (PSB), which consists of four separate synchrotrons. In the PSB the protons are arranged into bunches of order 10^{11} protons each. The proton bunches leave the PSB in groups of six with an energy of 1.4 GeV. The purpose of the PSB and its four synchrotrons is to increase the rate at which the next step in the chain, the proton synchrotron (PS), can be

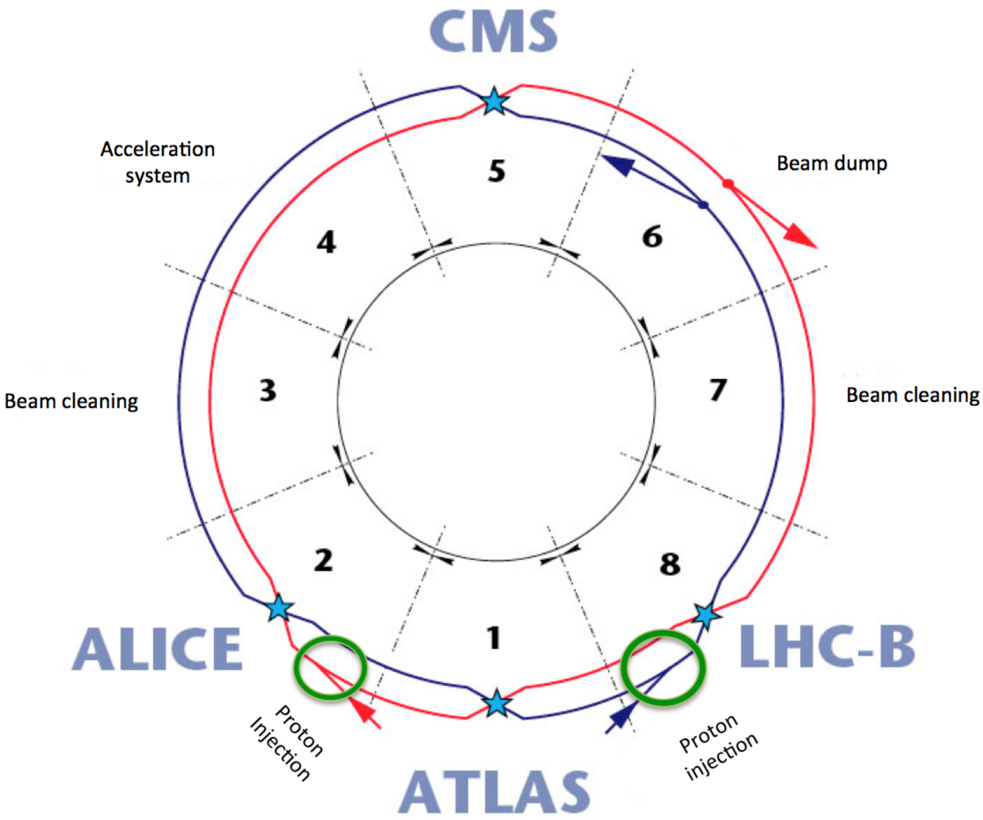


Figure 3.2: Schematic of the LHC collider showing the four main experiments, the eight numbered sections and an exaggerated layout of the curved and straight beam trajectories [45].

filled. The PS splits each of the six PSB bunches into 12 using radio frequency (RF) cavities of different frequency, then accelerates them to 25 GeV with a bunch spacing time of 25 ns (50 ns in Run 1). The bunches are then injected into the super proton synchrotron (SPS). The total number of bunches that can fit in the PS is 84, thus there are 72 bunches and 12 empty "spaces" in the PS. This gap is necessary to provide the kicker magnet for SPS injection with 300 ns of rise time. The SPS accelerates the bunches to 450 GeV before injecting them into the transfer lines to the LHC. The LHC is capable of holding 2,080 bunches in each beam at the design bunch spacing.

The LHC ring is divided into eight sections as shown in Fig. 3.2. There are four curved sections of the machine and four straight sections where the experiments are located. The LHC consists of two adjacent isolated beamlines for counter rotating beams of proton or heavy ion bunches. A common beamline is used for both beams only in the interaction regions where the collisions take place. Due to size limitations in the underground LHC tunnel, both beamlines are contained within the same mechanical structure. A "two-in-one" magnet design is necessary, where two sets of coils are arranged to produce flux in opposite directions within the beamlines. The magnet coils are made of a Niobium-titanium (NbTi) alloy and become superconducting below about 10K, but are operated at a temperature of 1.9K to give a field strength of 8.33T. The machine consists mostly of magnets with different numbers of poles, each of which has a different function. There are 1,232 dipole bending magnet sections, each 15 m in length, that constrain the beams in a circular trajectory. The magnets themselves have a slight longitudinal curvature, except those in the straight sections. Quadrupole magnets are used to squeeze and focus the beams, while higher order magnets and collimators are used to clean the beam and further focus it. Each beam is accelerated by eight 400 MHz RF cavities located in section 4, which provide 485 keV to each beam per orbit. Kicker magnets are used to quickly change the beam direction when injecting it from the SPS transfer lines or moving it towards the beam dump.

3.2 LHCb detector

The LHCb detector [46, 47] is a forward arm spectrometer covering the pseudo-rapidity region $1.6 < \eta < 4.9$, where $\eta = -\ln(\tan(\theta)/2)$ and θ is the polar angle from the beamline with a range of approximately $10 - 300$ mrad horizontally. In the vertical direction the angular acceptance is in the range $10 - 250$ mrad approximately. The main purpose of LHCb is to search for indirect evidence of physics beyond the SM and to measure *CP* violation in the heavy flavour sectors, specifically particles that contain b or c quarks such as the B and D mesons. The detector is designed to cover a narrow range in the forward region as b-$\bar{b}$ quark pairs are predominantly produced in the same forward or backward cone along the beam direction, due to the high probability of the interacting partons having unequal energies. The geometry of LHCb maximises the detector length within the available cavern space.

Decays of B and D mesons typically occur at secondary decay vertices (SVs) that are displaced from the primary interaction vertex (PV). This is due to the relatively long lifetimes of these mesons, $(1.520 \pm 0.004\,\text{ps})$ and $(0.4101 \pm 0.0015\,\text{ps})$ [9], for B^0 and D^0 respectively, which allow them to travel a few millimeters before decaying. Good spacial resolution of vertices is essential as many analyses, including those presented in this thesis, depend on measuring decay times. This is primarily achieved by reducing the number of vertices in any one event by operating the detector at a lower luminosity than that provided by the LHC. By offsetting the colliding beams by $1-3\sigma$ of the beam profile, a nominal luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ is maintained so that the number of PVs in each event is close to unity. The uncertainties associated with assigning the correct PV to an SV are greatly reduced by this modification.

Measurements of the luminosity at LHCb use two different methods [48]. The "van der Meer scan" method moves the beams in the two transverse directions to change the size of the overlapping region while measuring decay rates. The "beam gas imaging" method injects Argon gas into the interaction region. Decay vertices are constructed from particles produced by the beam interacting with the gas. The distribution of the vertices gives a measurement of the size of the interaction region and the rate of interactions.

LHCb has a right-handed coordinate system centered in the VELO sub-detector, with the beamline from LHC section 7 to section 0 defining the positive z-axis. The x-axis is horizontal and points away from the LHC centre and the y-axis is vertical and points upwards. The detector is split into two halves, the A-side, in the positive x direction (Airport side) and the C-side in the negative x direction (LHC Centre side).

The LHCb magnet has a window-frame design through which the beam passes. The complex shape of the magnet follows the outer edge of the detector acceptance and supplies an integrated magnetic field of 4 Tm for tracks 10 m in length. The field covers only a small region along the z-axis of the detector with uniform field lines pointing in the negative y-axis direction. The section of the detector on the interaction region side of the magnet is referred to as "downstream" of the magnet, while the opposite section is referred to as "upstream" of the magnet.

Like all particle physics detectors, LHCb is made up of different layers, shown in Fig. 3.3, that have different functions and measure different quantities. Three main aspects of the detector are covered here along with the associated sub-detectors that comprise the different layers. These are the tracking systems, which are comprised of the VELO, silicon and outer tracker sub-detectors; the particle identification system which consists of two ring imaging Cherenkov detectors, calorimeters and muon stations; and the trigger system which is implemented in hardware and software.

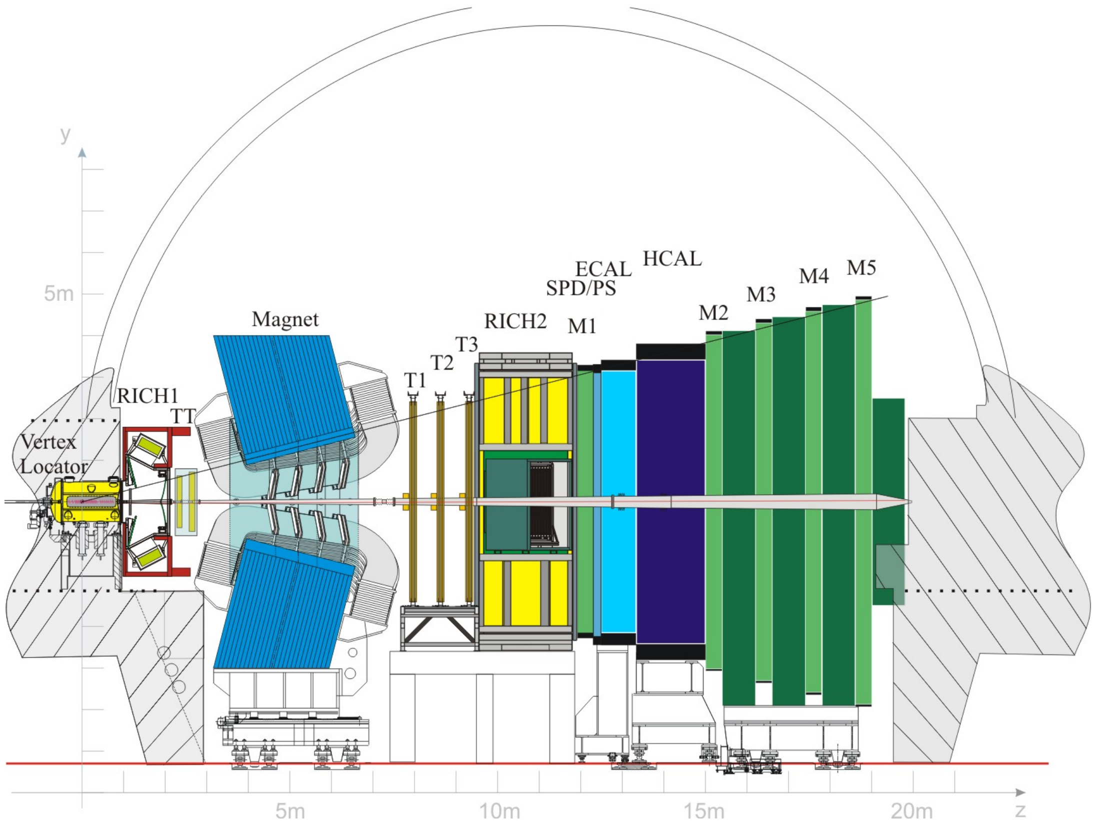


Figure 3.3: Cross-section of the LHCb detector. The beamline can be seen going from left to right through the centre of the apparatus. From left to right the sub detectors are the vertex locator (VELO), first ring imaging Cherenkov detector (RICH1), tracker Turicensis (TT), magnet, inner silicon trackers and outer trackers (T1 - T3), RICH2, scintillator pad detector and preshower (SPD/PS), electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL) and the muon stations (M1 - M5). The proton-proton collisions happen to the far left of the diagram, inside the VELO. The upper acceptance region is shown by a straight black line originating at the interaction point in the VELO. The positive x-axis points into the page [46].

3.2.1 Tracking

The particle tracking system at LHCb consists of four sub-detector elements. The vertex locator (VELO) surrounds the interaction region and provides precision measurements of PVs and SVs. The tracker Turicensis (TT) is located before the magnet and provides hit position information in that region. The outer tracker (T1-T3) modules are placed after the magnet and along with the hit positions before the magnet allow the curvature of a charged particle's track in the magnetic field to be determined. Each inner tracker (IT) module is placed in front of each

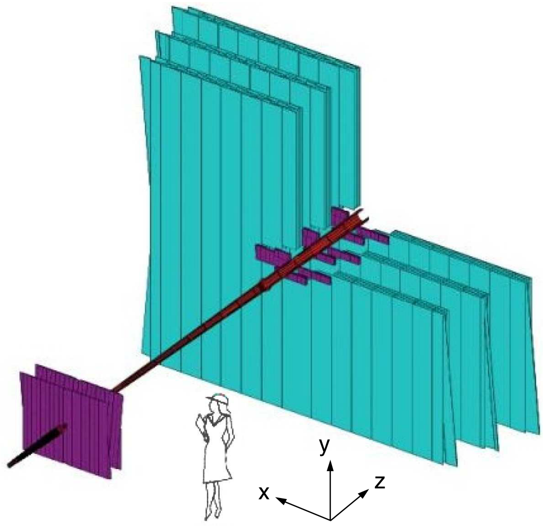


Figure 3.4: Diagram showing the LHCb tracking stations with the other sub-detectors removed and the beamline passing through the middle. The TT is the left most station shown in purple, the tracking stations (T1 - T3) are shown in teal with one quadrant cut away. The ITs are between the tracking stations and shown in purple also [46].

outer tracker module. The IT provides precision hit position measurements in the high occupancy region at the centre of the LHCb acceptance surrounding the beamline. The TT and IT are together known as the silicon tracker. Fig. 3.4 shows the position of the TT, IT and outer tracker sub-detectors.

Vertex Locator

This section describes the VELO sub-detector [49], which provides excellent vertex resolution that is essential to the physics goals of LHCb. The VELO contains 84 silicon strip sensors placed perpendicular to the beamline and arranged into modules that each contain 2 sensors. The sensors consist of an n-type silicon bulk with narrow strips of n^+-type silicon on the front surface. The strips have a higher dopant concentration than the bulk and thus more active charge carriers. A plane of p-type silicon is on the back surface of the bulk and a bias voltage is applied across the whole sensor. When a particle passes through the depletion zone, positive and negative charge carriers are produced. The negative charge carriers drift along the electric field to the n-type strips on the front surface of the bulk and the positive charge carriers drift to the sensor back plane. The distance between the strips, known as the "pitch" is a key parameter in defining the resolution of a sensor.

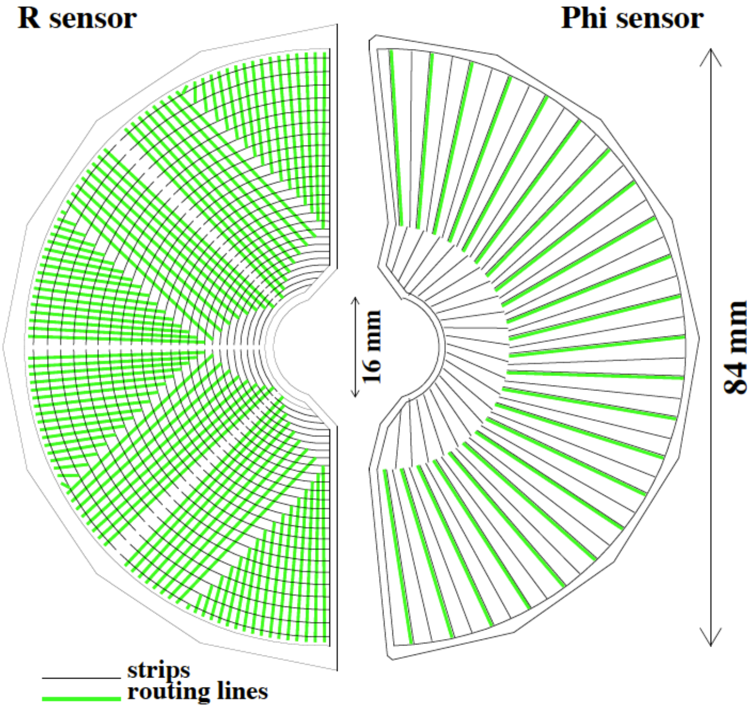


Figure 3.5: Diagram of an R-sensor and a ϕ-sensor, one from each half of the detector and thus from two different modules. The diagram shows the layout of the strips, as discussed in the text, and the routing lines used to readout the strips. The beam passes through the cut-out between the sensors.

Cylindrical coordinates, (R, ϕ, z) are used locally within the VELO, where z lies along the beamline, R is the radius from the beamline and ϕ is the azimuthal angle. The silicon strip sensors are arranged to measure positions in these coordinates. The sensors are semi-circular with a radius of 41.9 mm and a cut-out in the middle, through which the beam passes, as in Fig. 3.5. There are two sensor types, one for R coordinate measurements and one for ϕ coordinate measurements. R-sensors have their strips arranged in a circular pattern such that the strips near the beam are shorter. To reduce occupancy in any one strip each semi-circle is divided into four regions each covering 45° in ϕ so that each strip is divided into four smaller strips that are read out separately. The strip pitch near the beamline is smaller than the pitch at the outer edge of the sensor, which makes the resolution better at smaller values of R. The ϕ-sensors have their strips arranged along the radial direction. Again, to decrease occupancy, these sensors are split into two sections, one with $R < 17.25$ mm, which has a smaller pitch than the second section with $R > 17.25$ mm. An improved pattern recognition results from the introduction

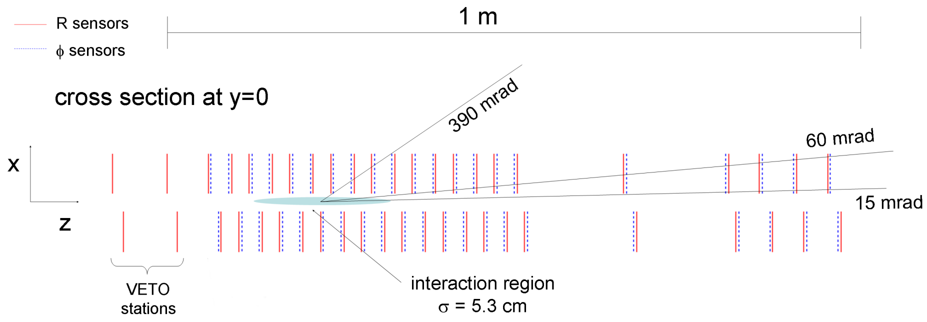


Figure 3.6: Diagram showing the layout of the VELO stations as described in the text along with relevant measurements [46].

of a small skew so that the strips are arranged at a small angle to the radial direction. The ϕ-sensors in alternate modules are flipped such that the sign of this skew changes. A module consists of one R-sensor and one ϕ-sensor mounted back-to-back on a support structure that also houses the electronic readout hybrid. As with most LHCb sub-detectors the VELO has two halves that separate along the beam position, but unlike other sub-detectors the VELO halves are regularly moved to centre them around the beamline. There are 21 stations each consisting of two modules mounted on opposite halves. The stations are arranged along the z direction with a finer spacing between modules in the region close to the interaction point, as in Fig. 3.6. The z coordinate of a sensor hit is known from the z position of the module and its uncertainty. It can be seen that the modules on opposite sides of the detector have slightly offset z positions. This is to allow for a small overlap between the two sides that precludes a gap in the ϕ coordinate coverage.

The LHC beam propagates through a highly evacuated (primary vacuum) beamline at a pressure better than 10^{-8} mbar [46]. The VELO is also operated in a vacuum (secondary vacuum) to reduce the material budget each particle must pass through. As the modules are so close to the beam during operation it was a great technical challenge to separate the two vacuum regions. This is achieved using a very thin aluminium sheet called the “RF foil”. Aluminium is used as it has a low Z number and is easily machined. The foil is 0.33 mm thick and must be corrugated so that it can fit between the stations’ sensors. It was thought that the corner of the sensor between the beamline cut-out and the straight edge might damage the thin foil. As a result the corner was softened as can be seen in Fig. 3.5, resulting in a small drop in overall acceptance.

During operation the sensors are 8 mm from the LHC beam. This makes them the closest active part of any LHC detector to the beam. However, during filling the beam position cannot be guaranteed. To protect the VELO it is necessary to retract the two halves including the RF foil by 29 mm. The two halves of the VELO can be remotely moved independently in the x-direction, and together in the y-direction, so that the VELO can centre around the interaction region. The procedure for closing the VELO is fully automatic and takes about 3 minutes. Vertices are constructed from the hits in the sensors, which indicate the position of the interaction region. The VELO halves move closer to this region, take more vertex measurements to better constrain the location of the interaction region, then move closer again.

To maintain the secondary vacuum and reduce radiation damage the sensors are cooled to $(-7 \pm 2)^{\circ}$C. The cooling substance is liquid CO_2, which transitions to a gas as it absorbs the heat in the sensors and electronics. It is supplied to the modules through 27 capillaries in each module. Low temperatures are used to reduce the mobility of defect ions that are separated from the silicon lattice, which is an effect of radiation damage. The sensors were operated with a reverse bias voltage of 150 V in Run 1, which is being increased as the sensors become more irradiated during Run 2.

The VELO pattern recognition [49] first groups R-sensor clusters consistent with a straight line from the interaction point then looks for ϕ-sensor hits along the same line. A candidate track is required to have at least three R and three ϕ-sensor hits. Another algorithm is then run to find clusters that can be linked with a candidate track that does not originate from the interaction point. A similar algorithm makes points from corresponding R and ϕ-sensors on the same module and uses them to make tracks.

The hit residual of the VELO is determined as the difference between the cluster position in a sensor and the intersection of the resulting track fit with that sensor. The correlation between the cluster position and the track fit is taken into account as the cluster is used in the track fit. A Gaussian fit to the distribution of residuals gives the hit resolution. This resolution increases linearly with the strip pitch as expected. Fig. 3.7(left) shows the resolution divided by the strip pitch against the projected track angle, defined as the angle between the track and the sensor normal, in the plane perpendicular to the sensor and to the normal plane of the strips. Tracks with larger projected angles have longer paths in the sensor and may cross more strips. The resolution is better for larger projected

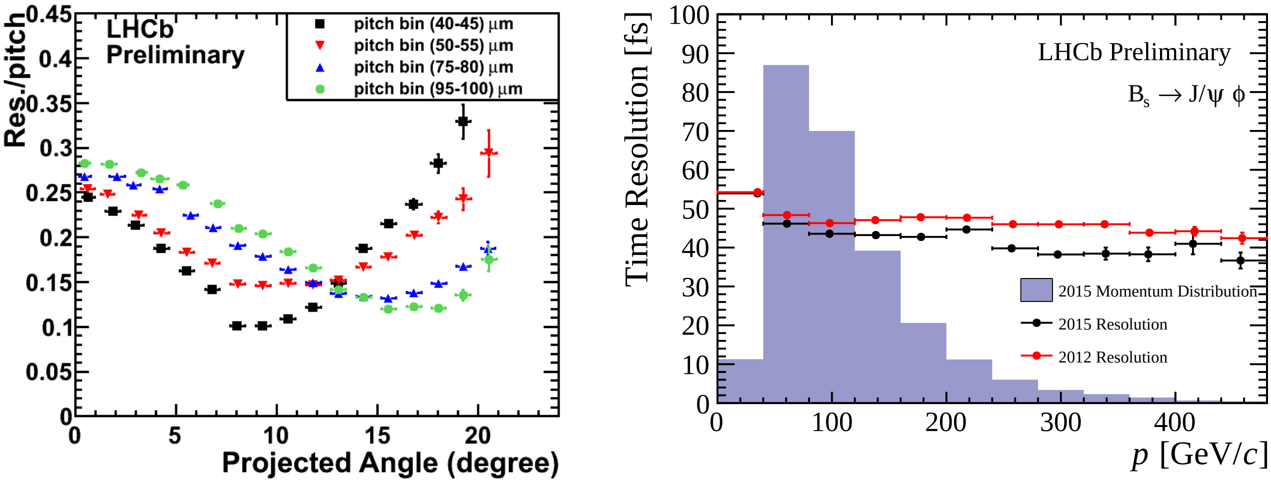


Figure 3.7: Shown are (left) the VELO hit resolution divided by strip pitch as a function of track projected angle and (right) the LHCb decay-time resolution against momentum, p, for 2012 and 2015 data taking [49].

angles as charge is more likely to be deposited in two strips and thus provide two measurements of the hit position. The weighted average of multiple strip clusters is taken as the hit position.

The PV resolution improves greatly with the number of tracks used to construct it and worsens by $5-10\%$ for each additional vertex in the event. The transverse PV resolution varies between 10 µm and 35 µm for between 5 and 40 tracks per PV. The impact parameter (IP) is the closest distance of approach between a vertex and a track. It is larger on average for tracks coming from a displaced SV, such as the daughter tracks from the decay vertex of a long-lived D^0 meson. The IP resolution depends strongly on $1/p_t$. It also depends on the VELO hit resolution, detector material effects, and crucially the distance between the PV and the first hit of the track in a VELO sensor. The decay-time resolution, shown in Fig. 3.7(right), is measured to be 50 fs in this four body decay, but has some dependence on the number of tracks in the vertex, the decay-length resolution, SV resolution and particle momentum.

Tracker Turicensis

The tracker Turicensis (“Zurich” tracker) [46, 47] is made from 500 µm p-on-n silicon strip sensors with a strip pitch of 183 µm and is placed downstream of the magnet. It consists of four layers of sensors each separated by 27 cm and covers the full LHCb acceptance region. The first and last sensor layers are arranged

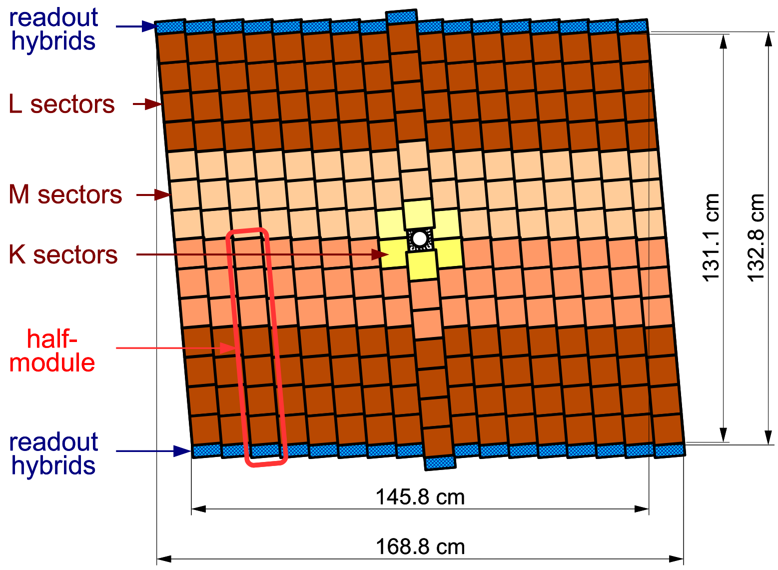


Figure 3.8: Diagram showing the second TT sensor layer with a half module indicated and the different readout sectors labelled [46].

vertically while the second and third have an angle to the vertical of $+5^{\circ}$ and -5°, respectively, which improves track finding. The first two layers are made from 34 half modules as indicated in Fig. 3.8, while the second two layers have four extra half modules to cover the larger acceptance. The half modules are made from seven sensors arranged into either two or three readout sectors such that the inner sectors with higher occupancy have fewer sensors to readout. The cooling, support and readout electronics for all sensors in each half module are located at the end of the module, outside the LHCb acceptance. Adjacent half modules are offset by about 1 cm in x and z to cover small acceptance gaps and to help with alignment calibration. The hit resolution of the TT was 53.5 µm with a hit efficiency of 99.8% in 2012.

Outer Tracker

The outer tracker [46, 50] is a drift-time detector made from 2.4 m long "straw-tubes", which consist of an anode wire inside a cathode cylinder with a diameter of 4.9 mm. An electric field is maintained between the anode and cathode and the cylinder is filled with gas. When a particle passes through the cylinder, the gas is ionised and the charge clouds drift to the electrodes. The choice of Argon, CO_2 and O_2 gases gives a drift time below 50 ns, which is necessary to reduce latency.

There are three stations in the outer tracker arranged along the z coordinate

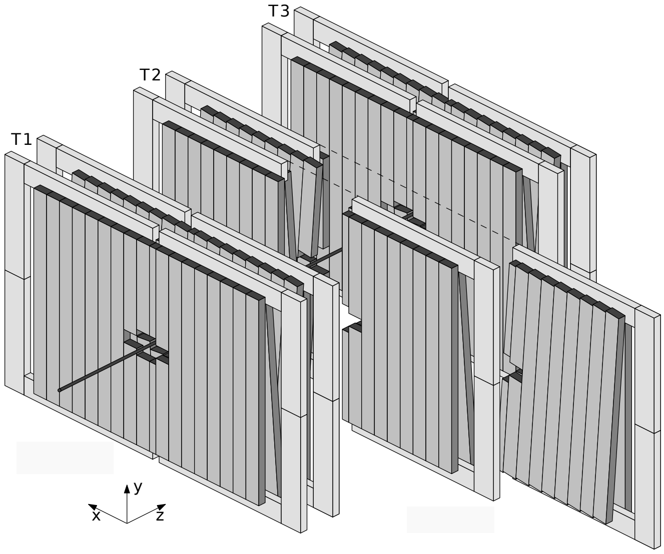


Figure 3.9: Diagram showing the three stations of the outer tracker, with the second station opened along the beamline. Each station has four layers shown in dark grey, each of which contains two layers of straw-tubes. The cross-shaped cut-out around the beamline is also shown [50].

direction. Each station has four layers, with the second and third offset from the vertical by $\pm 5^{\circ}$ as shown in Fig. 3.9. The layers are made from two monolayers of straw tubes that are horizontally offset from each other. There are 24 straw-tube layers in total. An inner cross-shaped section of the LHCb acceptance is not covered by the outer tracker. It is covered instead by the IT, which has a higher resolution.

The outer tracker hit efficiency from Run 1 data taking is found to be 99.2% and the hit resolution was approximately 200 μm.

Inner Tracker

The three IT [46,47] stations are placed in front of each of the outer tracker stations. Each consists of four isolated boxes, one above and one below the beamline and one to each side of the beamline, with small overlaps to remove acceptance gaps. The IT covers a region around the beamline that is approximately 40 cm high and 120 cm wide. Each box contains four layers of seven modules, and each module contains one or two sensors as shown in Fig. 3.10. The sensors are 7.6 cm × 11 cm and are silicon strips as in the TT. The single module sensors are 320 μm thick while the double module sensors are 410 μm thick, where all sensors have a strip pitch of 198 μm.

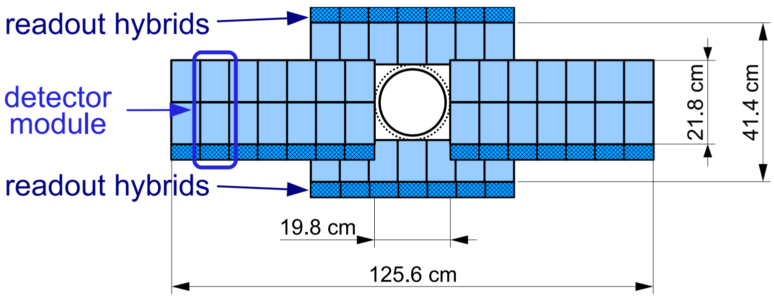


Figure 3.10: Diagram showing one detector layer of one IT station, each of the sections is from one of the four boxes. Each IT box contains four sensor layers. Sensors are shown as light blue rectangles. The upper and lower boxes have one sensor per module and the A and C-side boxes have two sensors per module. [46].

The hit efficiency of the IT in 2012 was 99.9% while the hit resolution was approximately 54.9 μm and varies with the track projected angle as the hit resolution depends on the number of strips crossed by the track.

3.2.2 Particle identification

RICH detectors

Ring imaging Cherenkov (RICH) detectors are used to identify different types of particles. They are essential for discriminating kaons from pions and for determining the final state particles of a decay. These detectors make use of the Cherenkov effect, which states that a charged particle with a velocity v moving through a medium with refractive index n such that $v > c/n$, where c is the speed of light in vacuum, will produce coherent radiation at a characteristic angle from the particle's trajectory. In other words, if a charged particle's velocity is greater than the local speed of light in that medium, coherent photons will be produced at a definite angle with respect to the particle trajectory. This angle is given by [51],

$$\cos\theta = \frac{1}{\beta n} = \frac{c\,m}{p\,n}, \tag{3.1}$$

where p and m are the particle momentum and mass, respectively. Thus the mass of an incident particle can be determined if the refractive index of the material and the particle momentum are known.

The ring imaging variant of these devices uses the cone of photons produced around a particle's trajectory. The photons are projected onto a plane of hybrid photon detectors (HPDs), where they appear as a ring. The particle's momentum

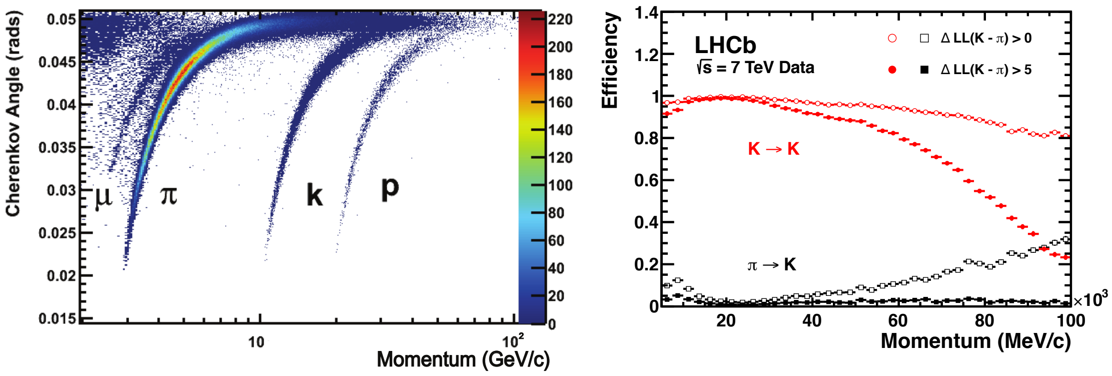


Figure 3.11: Graph of (left) momentum versus Cherenkov angle in the C_4F_{10} radiator for tracks with no overlapping Cherenkov rings, which shows the separation between different particles and (right) of kaon and pion identification efficiencies as kaons for 2011 data showing efficiencies for two likelihood hypotheses [52].

is determined from its curvature in the LHCb magnetic field, which is used along with various mass hypotheses to determine a set of Cherenkov angles. The mass hypothesis for which the Cherenkov angle fits best to the data is assigned to the particle with some likelihood.

In LHCb there are two RICH detectors, which are placed before and after the magnet and together cover a momentum range of $2-100$ GeV. The RICH devices perform better with low polar angle tracks. Thus RICH1 downstream of the magnet measures the particle velocity for low momentum tracks before they are deflected to high polar angles by the magnet, or swept out of the LHCb acceptance, and the upstream RICH2 measures the velocity of the high momentum tracks that are less deflected in the magnetic field. For this reason it is not necessary for RICH2 to cover the entire LHCb acceptance; only the angular acceptance of 120 mrad (100 mrad) horizontally(vertically) is covered.

RICH1 covers the entire LHCb acceptance except for a small region around the beamline. It contains silica aerogel radiator as well as C_4F_{10} gas radiator, the performance of which is shown in Fig. 3.11(left). The aerogel was removed after 2012 data taking due to its overall poor performance. RICH1 has one spherical mirror and two plane mirrors. The spherical mirror is placed at the upstream end and surrounds the beamline. The particles pass through the radiator as they enter from the VELO and the Cherenkov photons reflect off the spherical mirror. The incident particle then continues on through the spherical mirror as it is within the acceptance. For this reason the mirror cannot be made from glass as this would be a significant contribution to the radiation length. Instead a polymer is used that is supported by carbon fibre. The support structure of the mirror is mounted outside the acceptance. The planar mirrors reflect the Cherenkov photons parallelly onto the photon detectors. They are outside the acceptance so can be made of glass.

RICH2 performs a similar function to RICH1 except that it has a CF_4 gas radiator only. The photon detectors need to be outside the 300 mrad LHCb acceptance so the outer edges of the device are much larger.

To determine the type of a particle passing through the detector, a likelihood function, $\mathcal{L}$, is maximised as explained in [53]. It is not feasible to check all possible combinations of particle and particle type as there are a large number of possibilities. Instead, the pion mass hypothesis is used for all particles, as these are the most numerous. The mass hypothesis for each track is then changed in turn to either K, p or μ. The change of mass hypothesis for a single track that gives the largest increase in the likelihood is selected as its preferred hypothesis. This process is iterated on the remaining tracks to find the next largest increase in the likelihood. Using this process, the likelihood of a particle being of type x is given by the likelihood difference,

$$\Delta LL(x - \pi) := \ln\left(\frac{\mathcal{L}_x}{\mathcal{L}_\pi}\right). \tag{3.2}$$

The likelihood of a particle identified as a kaon is used frequently throughout this thesis, and is defined as PIDK $:= \Delta LL(K - \pi)$.

The efficiencies for identifying a kaon or pion as a kaon are shown in Fig. 3.11(right), which shows a trade-off between the efficiency for successfully identifying kaons and the misidentification efficiency for pions.

Calorimeters

The purpose of the LHCb calorimeters [46, 54, 55] is to detect the position and energy of charged and neutral particles and to provide some information on the type of particle detected. It must operate fast enough to provide transverse energy measurements to the trigger system every 4 µs. The LHCb calorimeters consist of four parts, in order along the z-direction these are the scintillator pad detector (SPD), the preshower (PS), the electronic calorimeter (ECAL) and the hadronic calorimeter (HCAL). All four parts have a similar structure and similar operation. The general structure of these detectors are alternating layers of absorbers and scintillator pads. Absorber pads are usually high Z materials that cause the particles passing through them to lose energy by producing secondary particles in a number of processes depending on the particle type. The secondary particles then enter the scintillator pads where charged particles deposit their energy in the form of light. The ECAL and HCAL calorimeters used in LHCb are of the "shashlik" type, which means that wavelength shifting fibers (WLS) are inserted through the detector perpendicular to the faces of the absorber and scintillator pads. These fibers receive the photons from the scintillator pads and emit a larger number of lower energy photons. The low energy photons are then guided through an optical cable to a photomultiplier tube (PMT) where the signals are amplified and counted. Shashlik type calorimeters are used as they allow the detector to be read out quickly.

The type of interaction in a calorimeter depends on the type of particle. Electrons lose most of their energy due to bremsstrahlung radiation while photons lose energy by pair production of electrons and positrons. An incident particle creates an electromagnetic shower in the calorimeter where secondary particles go on to produce tertiary particles and so on. This continues until the particles in the shower have an energy below the threshold needed to excite charged particles.

The SPD and PS calorimeters together are two scintillator pad detectors separated by 15 mm of lead absorbers. The purpose of these detectors is to remove a large background of low energy electrons before they reach the main electronic calorimeter.

The ECAL covers the entire LHCb acceptance except for a 25 mrad region around the beamline where there is high fluence. Fig. 3.12 shows one of the ECAL modules. The modules consist of 2 mm thick lead absorbers, 120 μm TYVEX paper and 4 mm scintillators.

Muon Stations

Detection of muons is essential to the physics goals of LHCb as a number of interesting B decays contain muons in their final states. The muon detection system [46, 47] comprises five muon stations as shown in Fig. 3.13. The muon stations are wire chambers that consist of a cathode plate and a number of anode wires separated by a gas. The gas is ionised when a charged particle passes through it, the ions are then accelerated by the electric field across the chamber and collected and counted by the anode wires. The LHCb muon stations contain Argon, CO_2 and CF_4.

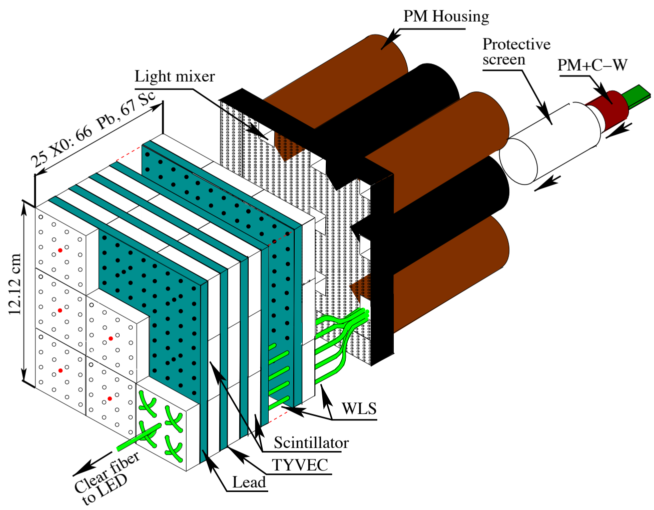


Figure 3.12: Diagram of an ECAL module from the inner region showing the absorber and scintillator layers. The wavelength shifting fibers in the shashlik design, and the photomultipliers are also shown [55].

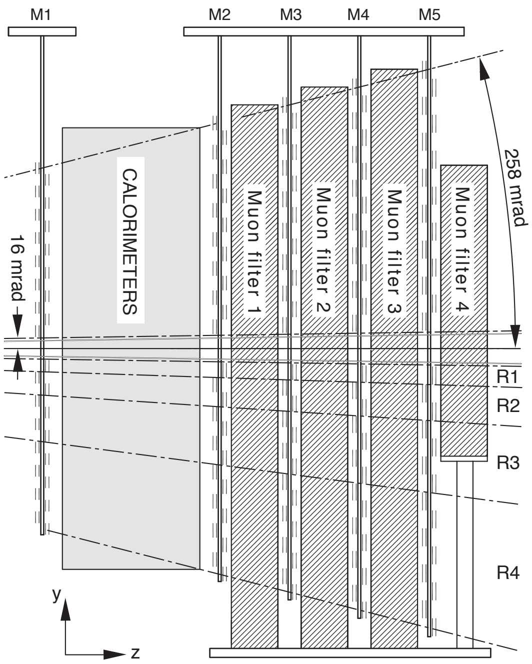


Figure 3.13: Diagram of the five muon stations showing their positions along the z direction and the angular size of the four sections of each station [46].

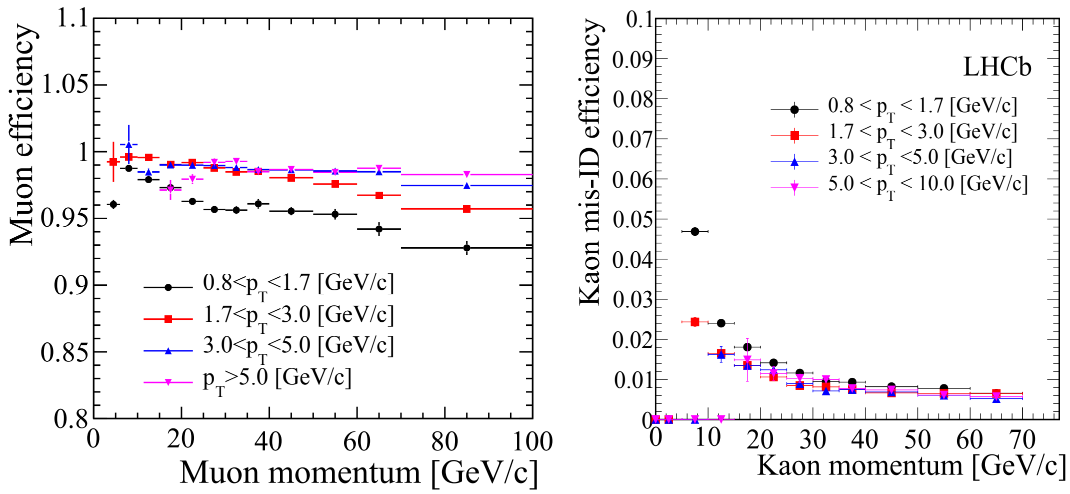


Figure 3.14: Graphs of (left) the muon identification efficiency and (right) the kaon misidentification efficiency, both of which are shown against particle momentum and shown for a range of particle p_t values [47].

One of the five muon stations is before the calorimeters to improve overall p_t measurements. The transverse displacement of a charged particle between M1 and M2 gives a measure of the p_t and having M1 further downstream improves resolution. M1 has a triple-GEM detector in its inner region as wire chambers would be damaged by the high fluence. Muon stations M2 – M5 are separated by 80 cm iron absorbers to reduce the energy of the incident muons and to stop particles that are not muons. The minimum momentum a muon would need to traverse these stations without being stopped is 6 GeV/c. If a particle traverses the full detector and has a hit in all 5 muon stations it is most likely a muon as other leptonic and hadronic particles are stopped in the calorimeters. The first three stations have good spatial resolution while the last two stations do not, as their purpose is only to determine if a particle had large penetrating power. The stations are divided into four different sections R1 - R4, which increase in size with distance from the beamline such that each section has similar occupancy. For a confirmed muon hit the candidate must be detected in M1 to M3. If the candidate has a momentum greater than 6 GeV/c it must also be seen in M4 and M5 to be considered as a muon. The reconstruction of the muon trajectory is done using the tracking stations and muon stations M2 and M3. The muon and kaon identification efficiencies as muons are shown in Fig. 3.14.

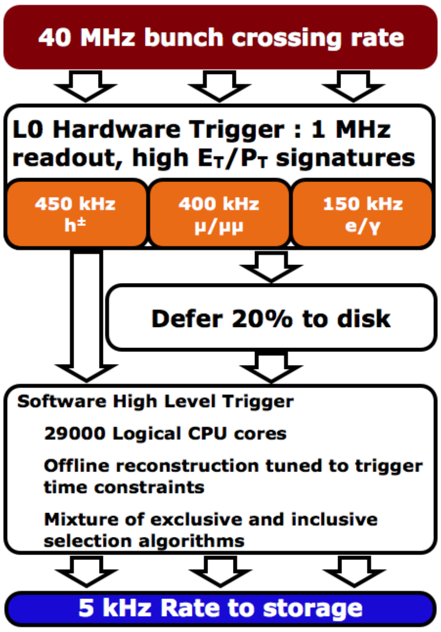


Figure 3.15: Schematic of the trigger data flow for Run 1. Note that the actual bunch crossing rate is lower than the design bunch crossing rate [58].

3.2.3 Trigger

The purpose of the LHCb trigger [46] is to reduce the data rate from the 20 MHz bunch crossing frequency (during Run 1) to a manageable 2 – 5 kHz storage frequency. It must have a high efficiency for selecting interesting physics events and a short latency. During Run 2 the bunch spacing time is 25 ns, but the number of bunches in the machine never reaches its theoretical maximum and the data rate is approximately 32 MHz. The collision frequency is also reduced by the lower luminosity delivered to LHCb.

The trigger is implemented in two stages, a hardware stage called Level 0 (L0) and a software stage known as the high level trigger (HLT). A schematic of the trigger data flow in Run 1 is shown in Fig. 3.15.

L0 trigger

The L0 trigger is implemented completely using custom built electronics that reduce the event rate to 1 MHz. The time in which the L0 trigger must decide to keep or discard an event is 2 μs. Not all sub-detectors are fast enough to be used in the L0 decision and none of them can supply all the available data. The sub-detectors supplying simplified data to the L0 decision unit are the calorimeters and the muon stations.

The calorimeter trigger starts by combining the energy of different hits into 2×2 clusters. It calculates the transverse energy, E_t, of these clusters and keeps only the hit with the highest energy for the L0 decision. Candidate particles are identified by the calorimeters as either photons, electrons or hadrons. If the number of hits in the SPD is greater than a threshold of order $600 - 900$ the event is vetoed due to the amount of time it would take to reconstruct. The total energy in the HCAL is summed to reject any events that appear to have no interactions.

The muon stations provide p_t and position information about candidate muons. The L0 trigger requires a muon candidate to have a hit in each of the five stations that can be joined by a trajectory through all hits to the interaction point. The two muon candidates with the highest p_t are kept. The reason for this is the importance of decays involving two muons to the physics goals of LHCb, for example the $B_s^0 \to \mu^+\mu^-$ decay [56].

The L0 threshold values for selecting an event based on the E_t in the calorimeters, the muon p_t and the number of SPD hits are given in references [47] and [57]. The L0 output contains one or all of the following measurements: the transverse momentum of two muons, the energy deposited in the calorimeter, and the calorimeter particle identification.

High level trigger (HLT)

The high level trigger is split into two stages, HLT1 and HLT2, which aim to remove uninteresting events. They are implemented entirely in software and as a result can be exactly simulated offline. HLT1 reduces the data rate from 1 MHz to 30 kHz at which point a full reconstruction can be done in HLT2. It begins by confirming L0 objects using a number of algorithms run in different sequences for different L0 decision. These algorithms involve either constructing candidate tracks in the VELO or in the other tracking stations (T-stations). If the L0 decision object was seen in the calorimeters it searches the T-stations in regions consistent with both positively and negatively charged particles. If the fit to the candidate's trajectory is not consistent with the L0 object the event is rejected.

HLT2 uses exclusive and inclusive "streams" that partially or fully reconstruct decays. Mother particles are constructed from the HLT1 candidates such that the combined mass lies within certain ranges. The details of the HLT2 triggers vary greatly between the various "trigger lines". The general idea is to select only composite particles that match the desired decay within certain loose selection requirements on mass and other observables. The HLT2 reduces the data rate from 30 kHz to the readout rate of 2-5 kHz. These events are then stored on disk to be analysed offline.

3.2.4 Trigger in Run 2

During data taking in Run 1 the trigger saved the full event information for each selected event to disk. Offline reconstruction and a further selection, known as "stripping", could then take place on the saved data. This strategy was necessary so that the saved events could be reconstructed with the most up-to-date alignment and calibration information of the detector and to give analysts access to the full event for offline calculations of efficiencies.

The maximum LHCb output rate is limited by computing challenges such as storage and processing time. To address these issues, the "turbo stream" [58] was introduced for Run 2. To reduce the storage space for each event, the turbo stream saves only the trigger objects associated to the decay of interest and discards the other tracks and measurements from the event. To minimize the number of data processing steps and thus reduce computing time, the stripping step is made redundant by the turbo stream. Events are buffered in this stream while the alignment and calibration of the detector are updated. The buffered events can then be processed with the updated calibration such that the data are of the same quality as the stripped data from Run 1.

This new stream is used along with the standard Run 1 streams, but it is expected that a larger number of analyses will switch to the turbo stream over time. Indeed, in future runs of the LHC it will not be possible to store and process events as was done in Run 1, due to the expected increase in the amount of data with the upgrade of the LHCb experiment in 2019/2020.

3.3 LHCb simulation and software

The data processing chain used by LHCb is described here, from initial collision data to user analyses. Many of the packages used on real data are also used to process simulated data. The common packages between simulation and real data are described as well as the packages and libraries used in the simulation.

The LHCb Monte Carlo (MC) simulation generates data-like events within a simulation of the LHCb environment. The simulated events may be used for determining selection efficiencies or acceptances, determining the fraction of a background within the signal sample, training machine learning algorithms, or determining the distributions of observables.

The first software package involved in the simulation is Gauss [59], which produces the particles and simulates their transport through the detector. Gauss relies on external packages for many applications. The Pythia 8 package [60] has been used since 2012 to simulate the *pp*-collisions with the desired centre-of-mass energy. It handles the initial parton-parton interactions as well as parton showers and the formation of complex hadrons. It uses the LHAPDF [61] set of parton distribution functions to model the internal structure of the colliding particles, with parameters tuned to match the multiplicity observed in LHCb real data. Next the EvtGen [62] package handles the evolution of the complex particles with time and their subsequent decays. A model of the LHCb detector is used along with the Geant4 [63] toolkit to model the traversal of the LHCb sub-detectors by the final state particles produced by EvtGen. The type of material and the type of particle are accounted for to determine, for example, the amount of charge deposited in a sensor by a single track.

Another LHCb software package, called Boole, takes the information supplied by Geant4 for each sensitive element in each sub-detector and simulates how the charge deposition is converted into digital signals. In silicon detectors this involves the distribution of charge along the particles track within the sensor and the diffusion of that charge to the sensor readouts.

The Moore package is responsible for running the three-stage trigger on real data and on simulated data. The L0 trigger is hardware based and so must be emulated by Moore. However, the HLT1 and HLT2 triggers are software based and can be run on the simulated data exactly as they are run on real data. From this point on the simulated data goes through the same processing chain as real data. The only significant difference is that the truth information about the simulated particles is stored. This includes the true particle ID, the true vertex positions, a unique ID for the true parent, and so on.

The next step in the data processing chain is the reconstruction of trigger objects and detector hits into tracks and vertices. This is done with the Brunel package. The reconstruction is run when the data is first taken, but can also be run after data taking to update the reconstruction with new versions of the software.

The final steps of the process are handled by the DaVinci package. The stripping step is run on the data as a further selection and reconstruction. It is similar to the HLT2 trigger, but has fewer time constraints and has access to particle identification (PID) information. Newer versions of the stripping are often run over the data to include new filtering algorithms and fix problems. In Run 2, data from the turbo stream is not processed by the stripping. The turbo stream has access to the full reconstruction and PID information and thus the stripping is not necessary. The objects in the turbo stream are reconstructed with a specially developed package called Tesla, which is a part of DaVinci.

At this point the analysts have access to the data files. The analysts use functionality within DaVinci to read the files and find events that have passed specific trigger and stripping lines. Typically, Root ntuples are then produced that can be easily manipulated for user analysis.

Part II

Measurement of the CP violation parameter A_Γ

Chapter 4

Overview

Described here is the analysis of two-body charm decays to extract the *CP* violation parameter A_Γ using the full LHCb dataset recorded in 2012. The analysis is based on that previously performed on 2010 data [64] and 2011 data [65] and is published in Physical Review Letters (PRL) [66] along with a complimentary method for measuring this observable [67].

The four tree-level two-body charm decays are [9],

- **Cabibbo favoured** (CF): $D^0 \to K^-\pi^+$ ($\mathcal{B} = (3.89 \pm 0.04)\%$).
- **Singly Cabibbo suppressed** (SCS): $D^0 \to K^+K^-$ ($\mathcal{B} = (3.97 \pm 0.07) \times 10^{-3}$) and $D^0 \to \pi^+\pi^-$ ($\mathcal{B} = (1.407 \pm 0.025) \times 10^{-3}$).
- **Doubly Cabibbo suppressed** (DCS): $D^0 \to K^+\pi^-$ ($\mathcal{B} = (1.385 \pm 0.027) \times 10^{-4}$).

Each decay also has an associated charge conjugate decay. Unless explicitly mentioned all statements implicitly apply to the respective charge conjugate decays. The tree level graphs for the four decays are shown in Fig. 4.1. The CF decay is called the right sign (RS) decay to distinguish it from the DCS, wrong sign (WS) decay. The final states of the SCS modes are *CP* eigenstates and thus can be used to measure A_Γ.

This analysis is based on measurements of the effective lifetimes of SCS decays as described in Section 2.5. Separate measurements are made of the D^0 and $\overline{D}^0$ lifetime parameters, which are then used to calculate A_Γ. The flavour of the D^0 meson is "tagged" by the strong decay $D^{*+} \to D^0\pi^+$, where the charge of the pion unambiguously gives the D^0 flavour. The pion is known as the "soft" pion, with symbol π_s. It typically has low momentum due to the small Q-value of the D^*

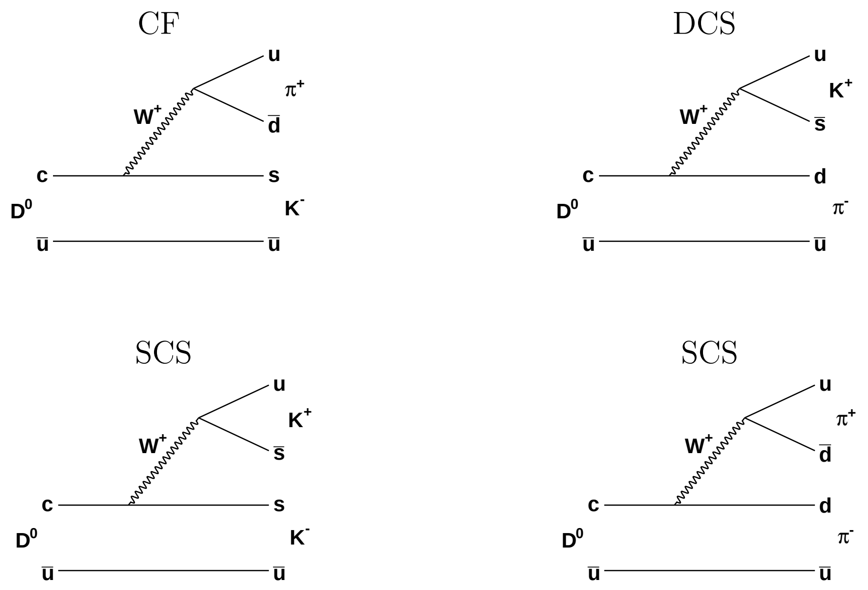


Figure 4.1: Tree level Feynman graphs of $D^0 \to h^+h^-$ decays. The Cabibbo favoured (CF) decay and the doubly Cabibbo suppressed (DCS) decays are shown on top and the two singly Cabibbo suppressed (SCS) decays to *CP* eigenstates are shown on the bottom.

decay, as the combined D^0 and π_s mass is similar to that of the D^*. It is possible for a random pion to be wrongly labelled as the soft pion, which may result in a mistagging of the D^0 flavour. This type of a background is known throughout as the random soft pion background, and is considered in more detail in Section 7.3. Using a strong decay for tagging guarantees that no flavour violating or *CP* violating process has occurred that could introduce a *CP* asymmetry between the two D^0 flavours. Only D^* candidates that have come directly (“promptly”) from the PV are used to measure A_Γ, as opposed to those resulting from B meson decays.

The relatively long lifetime of the D^0 meson means that it travels some distance from the PV before decaying. This displacement is used to distinguish D^0 daughter candidates from random combinations of short lived particles produced at the PV. Residual background of this type is called “combinatorial” background and is considered in Section 7.2. Other decay chains such as $B \to D^0X$ or $\Lambda_b \to (D^* \to D^0X')X$ are also possible and are referred to as “secondary” decays. These decays can be wrongly constructed as the decay of interest with a miscalculated decay length as illustrated in Fig. 4.2, which represents a small but irreducible background of potentially real D^0 candidates with a larger measured

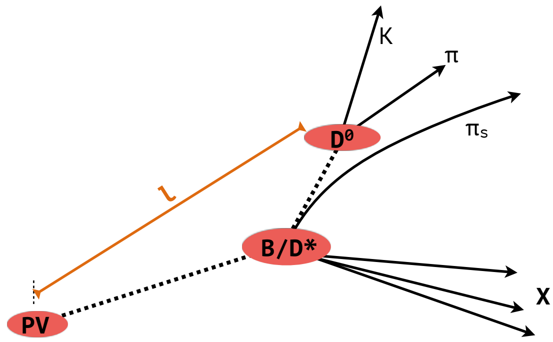


Figure 4.2: Sketch of a secondary D^0 decay chain of $B \to D^* X$ and $D^* \to D^0\pi$, showing the misreconstructed decay length, l, which results in a miscalculated D^0 decay time.

decay time. Secondary background and the observables used for distinguishing it from prompt signal decays are described in Section 8.1.

The analysis involves a data-driven method, described in Section 8.2, to correct for decay-time selection efficiencies ("acceptance") caused by the LHCb trigger selection criteria, geometry, and by offline selection criteria. A two stage fit procedure is employed to remove backgrounds and extract the prompt D^0 effective lifetime. The first 2D fit is to the D^0 mass, $m(D^0)$, and the mass difference between the D^0 and D^*, denoted as Δm, to determine the yield fractions of real correctly tagged D^0 decays, combinatorial decays and mistagged decays. The second fit is a 2D fit to the observable $\ln(\chi^2_{\rm IP}(D^0))$ and to the D^0 decay time in which the effective lifetime of the prompt signal component is the main fit parameter and the decay-time acceptances are accounted for. Here, $\chi^2_{\rm IP}$ is the difference in the χ^2 of the PV fit, with and without the D^0 candidate particle. This second fit allows prompt and secondary decays to be distinguished.

Chapter 5 describes the selection of the data and resulting biases to the D^0 decay time. Chapter 6 gives an overview of the maximum likelihood method used to fit the data and describes techniques used to model the various classes of decays, while Chapter 7 describes the first fit stage. In Chapter 8 the data-driven method for determining decay-time biases is described along with the fit to the D^0 decay time. Chapter 9 describes the systematic uncertainties and cross-checks while Chapter 10 summarises the results along with a combination of the results with those from previous analyses.

Chapter 5

Event selection

As discussed in Section 3.2.3, LHCb employs a series of hardware and software triggers that reduce the output data rate of the detector so that only those events that are of interest for physics analyses are stored. In order to select a sample of prompt two-body decays of D^0 mesons and to reduce backgrounds, selection requirements are applied to the measured observables.

5.1 Trigger selection

There are two categories of trigger decisions; a candidate particle is "trigger on signal" (TOS) if it is part of the reconstructed decay that passes this trigger. A candidate particle is "trigger independent of signal" (TIS) if it is not part of the reconstructed decay in the event that passes this trigger. This is a useful distinction as the TIS and TOS samples provide a data-driven method for determining trigger efficiencies as described in [68].

No specific requirements are made on the L0 trigger decisions of the event. Candidates are selected if they pass any of the L0 thresholds. In about 39% of the events, the L0 hadron trigger is fired by one of the two final-state hadrons of the D^0 decay. The remaining events are triggered by any other L0 hadron trigger in a way that is independent of the D^0 and its daughters (TIS).

At HLT1, candidates are required to be TOS with respect to the `Hlt1TrackAllL0` trigger line, which is described in [69]. This trigger is designed for finding B and D meson decays by looking for one high momentum daughter track originating from an SV. The selection thresholds are summarised in Table 5.1. The first step of the trigger is to reject events with high occupancy in the VELO, IT and outer tracker such that there are fewer than O(3,000) clusters in each sub-detector.

Table 5.1: Selection criteria for the `Hlt1TrackAllL0` trigger during final data taking in 2012. Changes to these thresholds throughout 2012 can be found in [57].

Observable	Selection requirement
VELO track hits	> 9
Missed VELO track hits	< 3
Track p	$> 10\,\mathrm{GeV}/c$
Track p_t	$> 1.7\,\mathrm{GeV}/c$
Track $\chi^2_{/\mathrm{ndf}}$	< 1.5
Track χ^2_{IP}	> 16

At the next step, hits in the VELO sensors are joined together to form various tracks, originating from the interaction region or displaced from the interaction region, as first described in Section 3.2.1. Not all candidate tracks can be considered due to time constraints. Only tracks with greater than nine hits and fewer than three "missed hits" are considered, where a missing hit is the absence of a hit in a sensor that is traversed by the candidate track. Selected tracks are extrapolated to "forward tracks" by searching the upstream tracking stations for hits in windows where the tracks might be found, given their trajectory before the magnet and considering both charge hypotheses. A Kalman fit is performed to the candidate tracks such that a selection can be made on the tracks' $\chi^2_{/\mathrm{ndf}}$ along with their p and p_t values. Requirements on the track χ^2_{IP} are made to remove backgrounds produced at the PV, mainly combinatorial background, and to favour events with displaced SVs. The p and p_t requirements aim to only select events with high momentum particles that are more likely to come from interesting high mass particles, such as a D^0 meson. The VELO hit requirements remove ghost tracks and ensure the selected tracks have good resolution. Two tracks passing these selection requirements are then considered to be D^0 daughter candidates.

At HLT2, the candidate D^0 daughters and their constructed mother particle, are required to be TOS with respect to one of the following trigger lines:

- `Hlt2CharmHadD02HH_D02`[hh],
- `Hlt2CharmHadD02HH_D02`[hh]`WideMass`,

where [hh] = {`kk`,`kpi`,`pipi`}, represents one of the three final states of interest.

Table 5.2: HLT2 trigger selection criteria at the end of 2012 data taking. Changes in these thresholds throughout 2012 can be found in [57]. DOCA is the distance of closest approach between two tracks, and DIRA is the cosine of the angle between the particle momentum vector and the direction vector connecting the PV to the SV ("direction angle").

Observable	Selection requirement
Both D^0 daughter momentum	$> 5.0\,\mathrm{GeV}/c$
Both D^0 daughter p_T	$> 800\,\mathrm{MeV}/c$
Largest D^0 daughter p_T	$> 1500\,\mathrm{MeV}/c$
Both daughter minimum χ^2_IP w.r.t. any PV	> 9.0
Both D^0 daughter track $\chi^2_{/\mathrm{ndf}}$	< 3.0
D^0 daughters' DOCA	$< 0.1\,\mathrm{mm}$
D^0 p_T	$> 2.0\,\mathrm{GeV}/c$
D^0 vertex fit $\chi^2_{/\mathrm{ndf}}$	< 10.0
D^0 PV flight distance $\chi^2_{/\mathrm{ndf}}$	> 40.0
D^0 PV DIRA	> 0.99985
D^0 mass, default line	$1790 < m(D^0) < 1950\,\mathrm{MeV}/c^2$
D^0 mass, wide mass line	$1715 < m(D^0) < 2085\,\mathrm{MeV}/c^2$
Number of tracks in event	< 180

The selection requirements made in these triggers are shown in Table 5.2.

The two D^0 daughter candidate tracks are combined to form a vertex with an associated fit $\chi^2_{/\mathrm{ndf}}$. To be considered further, the vertex fit of this new D^0 decay vertex must be less than 10. The mother D^0 candidate is constructed from the candidate daughters using the conservation of momentum, and assuming that there is no missing energy in the decay. This gives a momentum direction of travel for the mother particle and allows for selections to be made on the mother momentum and transverse momentum. The candidate D^0 particle's flight distance is determined by assuming it came from a PV and joining the two vertices with a line. Only D^0 candidates for which the cosine of the angle between the flight distance and the momentum direction (DIRA) is greater than 0.99985 are considered further. A mother D^0 does not leave hits in the detector as it doesn't travel far enough to reach the VELO sensors before decaying.

The combined mass of the two daughters and their momentum must lie in certain mass windows of the D^0 mass. The `WideMass` trigger lines have a wider

D^0 mass selection window and are scaled such that only 1/10 events are selected. These trigger lines are used to constrain the shape of combinatorial background in the D^0 mass window outside the range of the default triggers, which helps constrain the shape under the D^0 mass peak.

As in HLT1, momentum requirements favour events with a high mass parent and χ^2_{IP} requirements preferentially select daughter tracks coming from SVs. Selection requirements on the daughter χ^2_{IP}, the D^0 vertex and flight distance χ^2 and the DIRA, select candidate D^0 mesons that are likely to have come from the PV and are less likely to be secondary events. The mass windows chosen remove combinatorial background but leave the tails of the D^0 mass peak intact.

5.2 Stripping selection

An offline event selection known as the stripping is employed to further reduce background, improve the signal efficiency and to reduce computation time and storage space requirements by reducing the number of selected events. The candidate decays used in the analysis are required to pass the stripping line:

- `D2hhPromptDst2D2`[HH]`Line`,

where [HH] = {`RS`,`KK`,`PP`} represents one of the three final states. Recall that RS stands for right sign and represents the Cabibbo favoured $D^0 \to K^-\pi^+$ decay. The stripping line applies many of the same requirements as the trigger lines, as shown in Table 5.3, but has extra PID information from the RICH detectors, and reconstructs the tagging decay $D^{*+} \to D^0\pi^+$. The PID requirements applied to the D^0 daughters vary depending on the final state. For the $\pi^+\pi^-$ final state both daughters must have a PIDK < 0. For the $K^-\pi^+$ final state the kaon candidate must have a PIDK > 5, whereas the pion candidate must have a PIDK < 0. For the K^+K^- final state both daughters must have PIDK > 0 and at least one daughter must have PIDK > 5. The D^* is constructed by combining the D^0 candidate with positive and negative pions with track $\chi^2_{/\text{ndf}}$ values less than five. A candidate pair of a D^0 and a π_s is considered if their associated vertex fit $\chi^2_{/\text{ndf}}$ is less than 100 and their combined mass is consistent with being a $D^*(2010)$. Due to the low momentum of the π_s, the resolution on the D^* vertex is poor. For this reason, no further selection is made on the D^* vertex, and its position is not used in the separation of secondary decays.

Table 5.3: Stripping selection criteria, obtained from the stripping options script.

Observable	Selection requirement
Both D^0 daughter momentum	$> 5.0\,\mathrm{GeV}/c$
Both D^0 daughter p_T	$> 800\,\mathrm{MeV}/c$
Largest D^0 daughter p_T	$> 1500\,\mathrm{MeV}/c$
Both daughter minimum χ^2_IP w.r.t. any PV	> 9
Both D^0 daughter track $\chi^2_{/\mathrm{ndf}}$	< 3
D^0 daughters' DOCA	$< 0.07\,\mathrm{mm}$
Both D^0 daughter PIDK	See text
D^0 mass $(K^-\pi^+)$	$1765 < m(D^0) < 2065\,\mathrm{MeV}/c^2$
D^0 mass $(\pi^+\pi^-)$	$1790 < m(D^0) < 2065\,\mathrm{MeV}/c^2$
D^0 mass (K^+K^-)	$1765 < m(D^0) < 2065\,\mathrm{MeV}/c^2$
D^0 momentum	$> 5.0\,\mathrm{GeV}/c$
D^0 p_T	$> 2.0\,\mathrm{GeV}/c$
D^0 vertex $\chi^2_{/\mathrm{ndf}}$	< 10
D^0 flight distance χ^2	> 40
D^0 DIRA w.r.t. best PV	> 0.9999
$D^* - D^0$ mass difference	$< 160.0\,\mathrm{MeV}/c^2$
D^* vertex $\chi^2_{/\mathrm{ndf}}$	< 100.0
Soft pion track $\chi^2_{/\mathrm{ndf}}$	< 5.0

5.3 Further offline selection

Additional criteria, necessary for this analysis, are applied to the data after the stripping selection to remove additional backgrounds, reduce decay-time biases and simplify the fit method.

Events with multiple D^* candidates may result in, for example, the same D^0 candidate being used twice if it is associated with different soft pions. To avoid this eventuality, which could lead to a bias in the measured physics parameters, only one candidate is chosen at random and the remainder are rejected. Between 8% $(K^-\pi^+)$ and 13% $(\pi^+\pi^-, K^+K^-)$ of events have multiple candidates to which this selection is applied.

The D^0 decay time is required to lie in the range $0.25 - 10\,\mathrm{ps}$. The sample of candidates selected at very low decay-time values is prone to large uncertainties due to very low selection efficiencies and is removed to prevent instabilities in the fit method. At very high decay times, secondary decays become dominant and the number of prompt decays becomes negligible.

The decay radius of the D^0 must be less than 4 mm; this is the radial distance between the D^0 decay vertex and the PV. This requirement is applied in order to remove potential candidates produced by material interactions in the RF foil.

The number of "turning points" (see Section 8.2), which are used to determine the decay-time acceptance, is restricted to a maximum of two for each event. This selection removes very few events and greatly simplifies the inclusion of the acceptance in the fit PDF.

Further PID requirements are applied to the D^0 daughters to reduce backgrounds from misidentified daughter particles of D decays that peak or have tails in the D^0 mass window. These backgrounds are investigated further in Section 7.4. The applied PID requirements are final state dependent,

- $D^0 \to K^+K^-$: both kaons must have PIDK > 5,
- $D^0 \to K^-\pi^+$: the kaon must have PIDK > 5 and the pion must have PIDK < -5,
- $D^0 \to \pi^+\pi^-$: both pions must have PIDK < -5.

Candidates that pass the default trigger line are required to lie in the mass window [1826,1906] MeV/c^2. Candidates that pass the wide-mass trigger line are required to lie in one of the mass windows [1815,1826] MeV/c^2 or [1906,1915] MeV/c^2.

5.4 Decay-time biases

Unfortunately the selections described in previous sections have different efficiencies depending on the D^0 decay time. Here the selection criteria introducing these biases are discussed. The method to account for the biases is presented in Section 8.2.

Firstly the L0 trigger decision is examined. There is no obvious way in which any L0 trigger decision can cause a bias to the measured decay-time distribution. As a cross-check the sub-sample of events in which the D^0 is actively involved in the L0 trigger is analysed on its own without showing any inconsistencies.

A number of HLT1 selections have biasing effects on the D^0 decay time. The IP of a candidate daughter is related to the decay time, as the daughters of a long-lived particle will have larger IPs due to their displaced SV. Daughters of short-lived D^0 candidates will have smaller IPs as they originate closer to the PV. No selection is made on the IP, but requirements are made on χ^2_{IP}, which similarly biases the decay time. This requirement is necessary to remove combinatorial background events produced near the PV and attempts to select decays from an SV. A requirement on the number of VELO hits also biases the decay time as it favours short-lived particles. The VELO track finding algorithm, as described in Section 3.2.1, requires six sensor hits, while nine are required by the trigger. Long-lived particles travel further in the VELO and their daughters are more likely to exit the VELO before intersecting the required number of sensors.

The HLT2 stage of the trigger also has decay-time biasing selection requirements on χ^2_{IP}. Further biases at this stage are caused by requiring the D^0 and its daughters to have the same best primary vertex (BPV), defined as the PV for which the child's χ^2_{IP} is lowest. This condition reduces the number of kaons and pions that may be wrongly associated with a D^0. However, if an event has multiple interaction points it is more likely that a long lived D^0 will fail this requirement as it can move away from its own BPV and decay near the other interaction point. Thus the BPV assigned to the daughters could be different. A further requirement is made on the χ^2 of the D^0 flight distance. This variable is directly related to decay-time and will introduce a bias. The last relevant biasing requirement at the HLT2 stage is a requirement on the D^0 DIRA with respect to the PV. Candidates with shorter decay times are more likely to fail this selection due to larger direction angle resolution. The direction resolution between the PV and SV is larger for short-lived particles as the solid angle of probable directions is less constrained, resulting in a larger direction angle resolution.

Table 5.4: Number of candidates (n_{cand}) present for each decay after the stripping and after all selection criteria have been applied. The first two numbers in each column are the number of magnet polarity up and down events, respectively; the third number is their sum.

Decay	n_{cand} after stripping (Millions)	n_{cand} after offline cuts (Millions)
$D^0 \to \pi^+\pi^-$	$7.9 + 8.1 = 16.0$	$2.0 + 2.1 = 4.2$
$D^0 \to K^-\pi^+$	$60.9 + 59.9 = 120.8$	$37.6 + 39.4 = 77.0$
$D^0 \to K^+K^-$	$18.4 + 17.4 = 35.8$	$6.0 + 5.7 = 11.6$

The offline requirement to the D^0 radial flight distance introduces a decay-time bias. As D^0 candidates are created with transverse momentum and longer-lived candidates travel further, this requirement favours short-lived candidates with a low p_t.

5.5 Data sets

The number of candidates obtained after all the selections have been made are listed in Table 5.4. Each data sample is separated into disjoint blocks:

- Magnet polarity up or magnet polarity down;
- Before the first technical stop (runs 111,183 – 118,880), between technical stops (runs 119,956 – 128,492) or after the second technical stop (runs 129,534 – 133,785). Henceforth called run blocks one to three.

There are therefore six sub-datasets on which independent fits are performed. The motivation for this division is that the detector was opened during these technical stops. The separate blocks ensure that reconstruction differences between them are not obscured by combining them into one large sample. This is backed up by the fit results, as the reconstructed D^0 mass shape is different in the different samples. Each sub-dataset and the two flavour tags (D^0 or $\overline{D}^0$) are fitted independently, giving twelve fits for each decay.

Chapter 6

Parameter optimisation

The maximum likelihood estimate method, described and motivated in Section 6.1, is used to estimate the D^0 and $\overline{D}^0$ lifetime parameters from the datasets. Multiple discriminating observables from the datasets are used to separate the prompt signal from various background classes. The formalism of multidimensional probability density functions and their factorisation into smaller component parts is outlined in Section 6.2. The sPlot technique is described in Section 6.3, and further used in the technique presented in Section 6.4 for determining probability density functions of background classes without known distributions.

6.1 Maximum likelihood estimator

The measured value of an observable, x, for a single event, is a random number drawn from a distribution that encodes the probability of choosing a certain value of x. The probability density function (PDF), denoted by $f(x, \boldsymbol{\theta})$, is an analytic function that may depend on a set of parameters $\boldsymbol{\theta}$, which define the shape. A PDF is normalised such that the probability of making a measurement of any value is unity,

$$\int_{-\infty}^{\infty} f(x, \boldsymbol{\theta})\, \mathrm{d}x = 1. \tag{6.1}$$

Assuming no measurement bias, the distribution of x values is a consistent estimate of $f(x, \boldsymbol{\theta})$, meaning that in the limit of infinite events the measured and theoretical distributions are the same. Given a finite dataset of measurements, $\boldsymbol{x} = \{x_1, \ldots, x_N\}$ and assuming the underlying PDF, a method is needed to determine the parameters, $\boldsymbol{\theta}$, that best fit the data. A common method for determining PDF parameters from a finite dataset is the maximum likelihood estimator (MLE). The likelihood function is defined as the joint probability density of obtaining these measurements for some set of parameters,

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i, \boldsymbol{\theta}). \tag{6.2}$$

The parameters $\hat{\boldsymbol{\theta}}$ that maximise this function give the "best fit" to the data. In practice it is easier to maximise the natural logarithm of the likelihood function as the product of PDFs is then a sum. The global maximum of $\ln(\mathcal{L}(\boldsymbol{x}, \boldsymbol{\theta}))$ can be found by solving,

$$\frac{\partial \ln\left(\mathcal{L}(\boldsymbol{x}, \boldsymbol{\theta})\right)}{\partial \theta_a}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \tag{6.3}$$

for each of the parameters θ_a. In most cases a numerical approximation method is used to solve the series of equations. The standard errors on the parameters are related to the covariance matrix, which depends on the curvature of the log-likelihood function near the maximum,

$$\sigma_{\hat{\theta}_a} = \left(- \frac{\partial^2 \ln\left(\mathcal{L}(\boldsymbol{x}, \boldsymbol{\theta})\right)}{\partial \theta_a^2} \right)^{-1/2}. \tag{6.4}$$

Consider, for example, the true PDF for the exponential decay of a particle, $f(t, \tau) = (1/\tau)e^{-t/\tau}$, and a set of measured decay-time values $\{t_1, \ldots, t_N\}$. The derivative of the log-likelihood function is given by,

$$\frac{\partial \ln\left(\mathcal{L}(t_1, \ldots, t_N, \tau)\right)}{\partial \tau} = \sum_{i=1}^{N} \left(\frac{t_i}{\tau^2} - \frac{1}{\tau} \right). \tag{6.5}$$

Solving the differential equation gives the parameter value returning the maximum likelihood, $\hat{\tau} = (1/N)\sum t_i$ and standard error $\sigma_{\hat{\tau}} = \hat{\tau}/\sqrt{N}$.

The MLE is a widely used estimator due to its favourable properties [70], such as its invariance under parameter transformations. If a parameter value is transformed by some map, the ML estimate of the new parameter will simply be the transformation of the ML estimate of the original parameter. For infinite N the MLE reaches the minimum-variance bound, meaning any other estimator has the same or larger variance. It is not a consistent estimator for a small number of events, but is consistent in the limit that N tends to infinity. Similarly, it is an unbiased estimator in the limit of N tending to infinity, but has a bias of $1/N$ for a finite number of events.

With the total PDF constructed for this analysis the differential equations from Eq. (6.3) cannot be solved analytically. The MINUIT [71] program is used to maximise the log-likelihood by solving the set of differential equations numerically. The numerical solution uses the gradient of the log-likelihood to iteratively choose values of $\boldsymbol{\theta}$ that increase the log-likelihood until a maximum is found.

6.2 Multidimensional PDFs

To separate the prompt signal events from various "classes" of background events in the fits, a set $\boldsymbol{X}$ of discriminating observables is used. The PDF of each class for each observable must be constructed. The set of observables is not necessarily independent, thus a multi-dimensional PDF over $\boldsymbol{X}$ is required. Four lemmas for the manipulation of multi-dimensional PDFs are introduced here.

Lemma 1

The total fit PDF for all classes and the set of observables $\boldsymbol{X}$ is given by,

$$f(\boldsymbol{X}) = \sum_{class} P(class) f_{class}(\boldsymbol{X}), \tag{6.6}$$

where $P(class)$ is the fraction of events from this class relative to the total number of events, such that the total fraction is $\sum_{class} P(class) = 1$. The PDFs for each class are denoted by f_{class}. The relative fractions of the classes in the analysis are not known and are determined using an MLE.

Lemma 2

The probability that an event is from a specific class given the set of its measurements, $\boldsymbol{X}_i$, is given by,

$$P(class|\boldsymbol{X}_i) = \frac{P(class) f_{class}(\boldsymbol{X}_i)}{\sum_{class'} P(class') f_{class'}(\boldsymbol{X}_i)}, \tag{6.7}$$

which is simply the probability that the event is from this class, normalised with respect to its probability of being in all other classes. This lemma holds for all $\boldsymbol{X}_i$, and thus for $\boldsymbol{X}$.

Lemma 3

A multi-dimensional PDF with a set of observables $\boldsymbol{X} = \{x, y, z\}$, can be factorised as follows,

$$\begin{aligned} f_{class}(x, y, z) &= f_{class,x,y}(x, y|z)\ \ f_{class,z}(z) \\ &= f_{class,x}(x|y, z)\ \ f_{class,y}(y|z)\ \ f_{class,z}(z), \\ &:= f_{class}(x|y, z)\ \ f_{class}(y|z)\ \ f_{class}(z), \end{aligned} \tag{6.8}$$

where the subscripts of f are simply labels to distinguish between different PDFs and have been dropped for brevity in the final line. In what follows, it is assumed that PDFs with different arguments represent different mathematical functions. In these expressions, a PDF such as $f_{class}(z)$ is the one dimensional distribution of the observable z, which does not depend on the distributions of x and y. Note that z is not a special observable, any one of the variables could have been chosen as the first independent observable. Thus the two factorisations are equivalent,

$$f_{class}(y|z)f_{class}(z) = f_{class}(z|y)f_{class}(y). \tag{6.9}$$

The conditional PDF, $f_{class}(y|z)$, is the one dimensional PDF of the observable y in which the distribution is correlated to the z distribution. It is the projection of the full multidimensional PDF onto the y axis and integrated over the z observable distribution. If some of these parameters are independent of each other, y and z say, then,

$$f_{class}(y|z) = f_{class}(y). \tag{6.10}$$

Lemma 4

If $\boldsymbol{X}$ can be written as two independent sets of observables, $\boldsymbol{X} = \boldsymbol{Y} \cup \boldsymbol{Z}$, then the total PDF can be written in two parts by using the previous lemmas,

$$\begin{aligned}
f(\boldsymbol{Y} \cup \boldsymbol{Z}) &= \sum_{class} f_{class}(\boldsymbol{Y}) f_{class}(\boldsymbol{Z}) P(class), \\
&= \left[\sum_{class} f_{class}(\boldsymbol{Y}) f_{class}(\boldsymbol{Z}) P(class)\right] \sum_{class} f_{class}(\boldsymbol{Y}) P(class) \frac{1}{\sum_{class} f_{class}(\boldsymbol{Y}) P(class)}, \\
&= \left[\sum_{class} \frac{f_{class}(\boldsymbol{Y}) f_{class}(\boldsymbol{Z}) P(class)}{\sum_{class'} f_{class'}(\boldsymbol{Y}) P(class')}\right] \sum_{class} f_{class}(\boldsymbol{Y}) P(class), \\
&= \left[\sum_{class} f_{class}(\boldsymbol{Z}) \frac{f_{class}(\boldsymbol{Y}) P(class)}{\sum_{class'} f_{class'}(\boldsymbol{Y}) P(class')}\right] \sum_{class} f_{class}(\boldsymbol{Y}) P(class), \\
&= \left[\sum_{class} f_{class}(\boldsymbol{Z}) P(class|\boldsymbol{Y})\right] \left[\sum_{class} f_{class}(\boldsymbol{Y}) P(class)\right].
\end{aligned}$$

The log-likelihood function constructed from this PDF is the sum of two terms, one of which (the $\boldsymbol{Y}$ term in this case) is independent of the other term. The independent term can be considered separately as a single log-likelihood for which the maximising parameters can be determined. The optimum parameters can then be fixed in a second MLE using the second term in the full log-likelihood function. The CPU time required for the execution of an MLE depends on the size of the parameter space to be searched. Each extra parameter adds an extra axis to this space and thus the CPU time has an exponential dependence on the number of parameters. Factorising the fit PDF into an independent part and a dependent part by finding two sets of independent observables significantly reduces the computation time.

6.3 sPlots

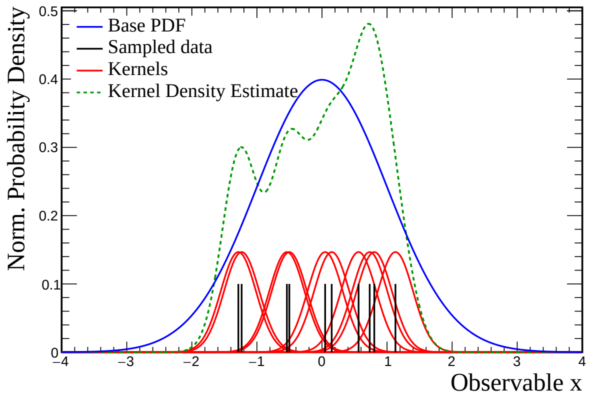
Presented here is a brief account of the sPlot technique, which is described more fully in [72]. In the case where the distribution of an observable x for some class of decays is unknown, the sPlot technique allows this distribution to be determined, even in the presence of significant backgrounds from other classes of decays. For this technique, a discriminating observable m is necessary to first distinguish the various decay classes. A maximum likelihood fit to the observable m gives the covariance matrix V, and the sWeight for a single event is given by,

$$ {}_s\mathcal{P}_n(m) := \frac{\sum_{class} V_{n,class} f_{class}(m)}{\sum_{class} N_{class} f_{class}(m)}, \tag{6.11} $$

where the sums are over the various classes and N_{class} is the expected number of events from a certain class. The function $f_{class}(m)$ is the distribution of a certain class in the discriminating observable m. The sWeights are thus the covariance weighted probability of a single decay belonging to a certain class of decays. By weighting each decay with its corresponding sWeight and plotting the weighted distribution of x, the contribution from the desired class is revealed, and the contributions from the other classes are removed. The sPlot technique assumes that the discriminating observable, m and the control observable x are uncorrelated.

6.4 Kernel density estimation

The MLE relies on some prior knowledge of the PDFs for each class of events. A number of the backgrounds in this analysis have unknown distributions in the discriminating observables and in decay time. It is necessary to choose a continuous function to describe the PDF of these background components in the MLE. Given a set of data drawn from an unknown PDF, the method described here produces a continuous function that estimates the true distribution of the unknown PDF.



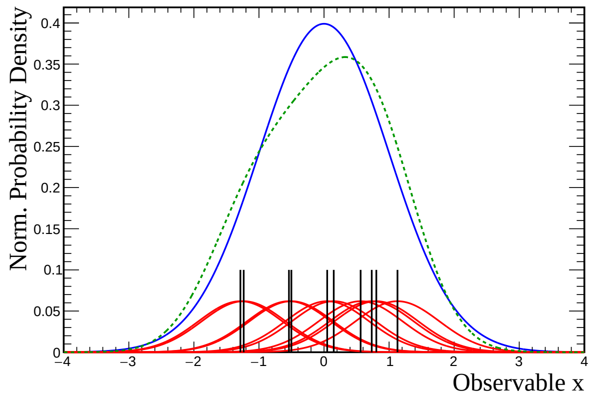


Figure 6.1: Examples of kernel density estimation. A total of ten event data points (blank lines) are drawn from a Gaussian distribution (blue) of mean zero and width equal to one. A Gaussian kernel (red) is applied to each data point and the kernel density estimate (dashed green) is constructed from them. Shown is the kernel density estimate, using (left) the initial bandwidth estimate and (right) the optimum bandwidth after iterating as described in the text. A reasonable agreement is seen between the true distribution and the kernel density estimate.

Kernel density estimation assigns a “kernel”, $k(x)$, to each event in the set of data. Any function can be used as the kernel, but in this case a Gaussian with width σ is used. The mean of each Gaussian kernel is centered on the value of the observable x, for each event. The width is known as the “bandwidth” and can be assigned any value, but its optimum value must be determined. The individual kernels can be summed to form a continuous distribution, $K(x)$, that estimates the true PDF. An initial estimate of the bandwidth is given by [73],

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}}, \tag{6.12}$$

where $\hat{\sigma}$ is the standard deviation of the events in the dataset and n is the number of events. As can be seen in Fig. 6.1(left) for a true Gaussian PDF, this bandwidth tends to be too small, resulting in an estimate that changes too quickly across the observable.

The optimum bandwidth is given by [73],

$$h^* = \left(\frac{R(K)}{\sigma_K^4 R(f'')n}\right)^{\frac{1}{5}}, \tag{6.13}$$

where σ_K is the variance of the summed kernel functions, f is the true distribution of the data and R is a functor that encodes the "roughness" of its argument, given by,

$$R(f) = \int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x. \tag{6.14}$$

Evidently, this optimum bandwidth cannot be calculated without knowledge of the underlying distribution, f. The first estimate of $K(x)$ with the bandwidth of Eq. (6.12) is used to approximate f. The approximation can be further improved by iteratively calculating the optimum bandwidth using successive $K(x)$ functions. At each step, $K(x)$ from the previous step is used to calculate the new optimum bandwidth and thus the next $K(x)$. Fig. 6.1(right) shows the kernel density estimate calculated in this way for 10 events. The agreement between the resulting and the original distributions improves with the addition of more events, however the agreement with only 10 events is good.

The optimum bandwidth may vary across the observable. To account for this, the bandwidth can be parameterised in terms of x. For example, if the bandwidth was proportional to the local standard deviation about each x value, the kernels in the Gaussian tail would be wider, and would better approximate the true distribution. However, in this analysis all kernels are fixed to have the same bandwidth.

In the case already described, it was known that the values of x for each event were drawn from the same PDF and thus are events of the same class. In the more common situation, the dataset contains a mixture of different classes that can be separated in some discriminating observable m. To determine the kernel density estimate of one of these classes with an unknown PDF, they must first be separated out from the other events. This is done in the analysis using the sPlot technique, described in Section 6.3.

In this analysis the kernel density estimation technique is applied to the total data, sWeighted for a certain class, to produce a PDF of that class in a chosen control observable.

Chapter 7

Fit to mass variables

A description of the maximum likelihood fit used to discriminate between signal and background events in the first fit stage of the analysis is presented in this chapter.

The discriminating observables used throughout the analysis are the D^0 invariant mass, $m(D^0)$, the invariant mass difference $\Delta m := m(D^*) - m(D^0)$, the proper decay time of the D^0, denoted t, and for secondary separation, $\ln(\chi^2_{\text{IP}}(D^0))$. The total PDF used for the analysis is,

$$f\big(m, \Delta m, \ln(\chi^2_{\text{IP}}), t\big) = \sum_{class} P(class) f_{class}(m, \Delta m, \ln(\chi^2_{\text{IP}}), t), \tag{7.1}$$

where superfluous arguments have been dropped for clarity. The factorisation of this PDF, which motivates splitting the fit into two stages, is described in Section 7.1.

The background components present in all three final states and their associated subclasses are shown in Table 7.1. Sources of background that contaminate the

Table 7.1: The classes and subclasses of the analysis that represent the different types and subtypes of $D^0 \to h^+h^-$ signal and background decays.

Class	**Subclass**
Signal	Prompt Secondary
Random soft pion background (π_s)	Prompt Secondary
Combinatorial background	-
Specific misreconstructed backgrounds	-

signal $D^0 \to h^+h^-$ decays in the first fit stage are described here in terms of their origins and their probability densities in discriminating observables. Backgrounds in the second "time" fit stage are described later in Chapter 8. Backgrounds specific to each of the three final states are described in Section 7.4. Example distributions are shown from samples of simulated data, the selection for which is the same as the real data, described in Chapter 5.

7.1 Factorisation of the full fit PDF

Using lemma 4 from Section 6.2, the full fit PDF can be factorised into one independent part and one dependent part,

$$f\big(m, \Delta m, \ln(\chi^2_{\mathrm{IP}}), t\big) = \left[\sum_{class} P(class) f_{class}(m, \Delta m)\right] \times \left[\sum_{class} P(class|m, \Delta m) f_{class}(\ln(\chi^2_{\mathrm{IP}}), t)\right]. \tag{7.2}$$

This motivates splitting the analysis method into two separate MLEs. The first fit to the "mass" observables $\{m, \Delta m\}$ supplies the $P(class)$ relative fractions. Then the probabilities $P(class|m, \Delta m)$ for events to be signal or background based on the distributions of $m(D^0)$ and Δm, can be calculated using lemma 2. With the known relative fractions of the classes, a fit to the "time" variables, $\{t, \ln(\chi^2_{\mathrm{IP}})\}$ can provide the fractions of decay subclasses and measure the D^0 decay time.

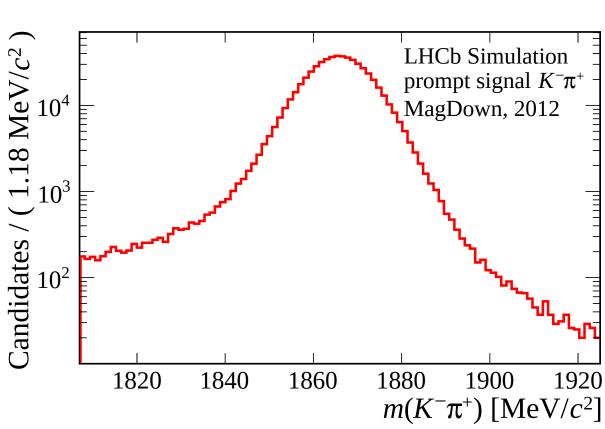
By factorising the total PDF into two parts, it is assumed that the two variable sets are uncorrelated. This is not exactly true for the mass and time set variables. An investigation of these correlations is presented in Section 9.4.3, along with a determination of their effect on the final result.

7.2 PDFs in D^0 mass

The invariant mass of a parent particle decaying to two daughters is calculated as:

$$\begin{aligned} m(D^0)^2 &= (E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2, \\ &= \left(\sqrt{m_1^2 + |\vec{p}_1|^2} + \sqrt{m_2^2 + |\vec{p}_2|^2}\right)^2 - |\vec{p}_1 + \vec{p}_2|^2, \end{aligned}$$

where $i = 1, 2$ denotes the daughter particle measurements. In the $m(D^0)$ fit, the signal is modelled as two Gaussian functions with the same mean ($\mu_{m,sig.}$) but



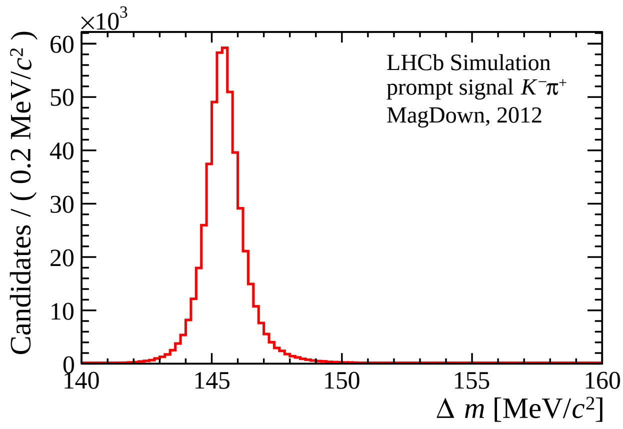


Figure 7.1: Distribution of simulated prompt signal events with magnet polarity down and 2012 conditions for (left) the D^0 invariant mass showing the radiative tail, (right) the D^* and D^0 invariant mass difference.

different widths ($\sigma_{m_i,sig.}$), thus giving three parameters plus one more to control the relative fraction between the Gaussians ($P(m_1|sig.)$). Pions in the D^0 final state can radiate photons as they travel through the detector, which reduces their momentum and creates a long tail on the low mass side of the $m(D^0)$ peak. To account for this, the second signal Gaussian is given a power law tail, known as a "Crystal ball" and given by [74],

$$f(x) = \begin{cases} \exp\left\{-\frac{1}{2}t^2\right\} & t \geq -|\alpha| \\ \frac{A}{(B-t)^n} & t < -|\alpha| \end{cases}, \tag{7.3}$$

where $t := \frac{\alpha}{|\alpha|} \cdot \frac{x-\mu}{\sigma}$, σ is the width of the distribution, α defines where the power-law tail begins, μ is the most probable value, $A := (n/|\alpha|)^n \cdot \exp\left\{-\frac{1}{2}|\alpha|^2\right\}$, $B := (n/|\alpha| - |\alpha|)$ and n is the power-law exponent. The form of the PDF presented here has not been normalised but an analytic integration is performed in the analysis framework. This increases the length of the signal tail on the low mass side of the peak. Fig. 7.1(left) shows the $m(D^0)$ distribution for simulated prompt signal events. The Crystal ball has two parameters ($\alpha_{sig.}$,$n_{sig.}$) and thus the signal PDF has six parameters in total.

No power tail is needed in the $D^0 \to K^+K^-$ final state as kaons radiate photons less often, thus there are only four parameters in total. Due to the smaller size of the $D^0 \to \pi^+\pi^-$ sample and the relatively large size of the background component, the Crystal ball parameters could not be allowed to float in the fit. Instead, a fit to simulated data is used to constrain the parameters as $\alpha_{sig.} = 2.05$ and $n_{sig.} = 1$. As the CF $D^0 \to K^-\pi^+$ sample is very large, the fit is more sensitive

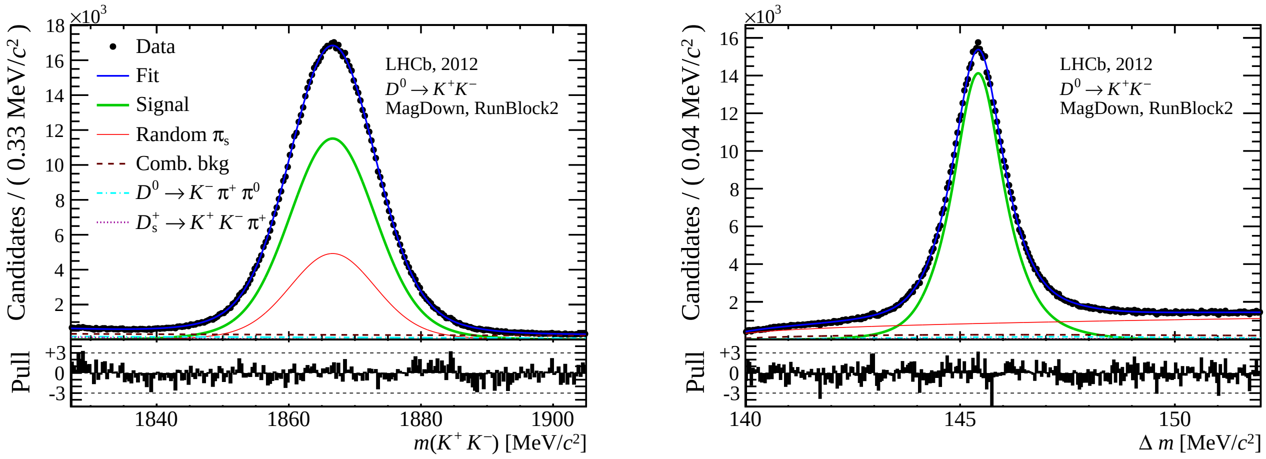


Figure 7.2: Projections of the fit to (left) D^0 mass and (right) Δm for the $D^0 \to K^+K^-$, magnet polarity down sub-dataset in the second run block.

to contributions to the mass peak that have different or varying resolutions. An extra Gaussian is needed to describe these resolutions, adding one more width and one more fraction, for a total of eight parameters.

The main purpose of a fit to $m(D^0)$ is to determine the relative fraction of combinatorial background events in the dataset. These are combinations of random particles produced at the primary interaction vertex and from partially reconstructed B decays. Combinatorial background from the primary vertex is expected to be short lived and to have a falling distribution in $m(D^0)$. The component from B decays is long-lived and is expected to have a flat or rising distribution in $m(D^0)$. The momentum direction of a misreconstructed two-body decay, originating from a three body decay, will not point back to the PV. As a result, requirements on the momentum direction greatly suppress combinatorial background. Vertex quality requirements, as in the selection requirement tables of Chapter 5, also suppress this background as the two daughter tracks may not originate from the same vertex. Combinatorial background is modelled in the fit with an exponential whose single decay parameter ($e_{comb.}$) is free to float.

The $m(D^0)$ projection of the 2D fit to $m(D^0)$ and Δm is shown in Fig. 7.2(left) for one K^+K^- sub-dataset. The fit parameters and values for this one sample are shown in Table 7.2. Detailed results for all final states can be found in Appendix B.

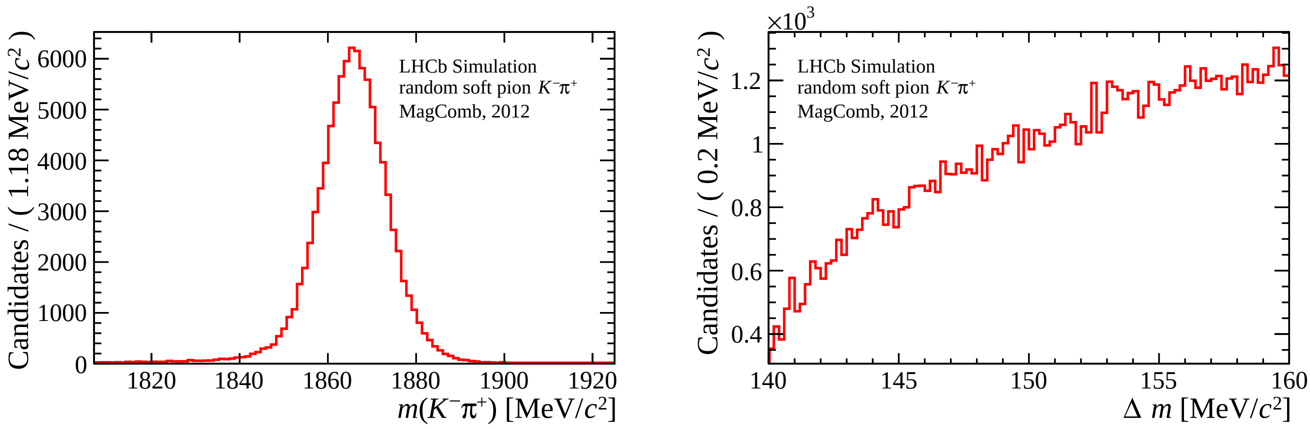


Figure 7.3: Distribution of simulated random soft pion events with magnet polarity up and down with 2012 conditions for (left) the D^0 invariant mass, which has the same distribution as correctly tagged signal events, (right) the D^* and D^0 invariant mass difference, which increases smoothly from the pion mass threshold.

7.3 PDFs in Δm

The Δm distribution for correctly tagged prompt signal decays from simulation is shown in Fig. 7.1(right). The peak position is related to the pion rest mass and the average momentum of the soft pion and D^0. It is possible for a D^* candidate to be formed by combining a D^0 meson and a random track that is incorrectly designated as the soft pion. This background is called the random soft pion background. Its $m(D^0)$ distribution is the same as that for correctly tagged signal as it contains a real D^0 decay to the correct final state. Thus identical PDFs are used to describe signal and random soft pion background in $m(D^0)$.

The sign of the charge on the random soft pion may not tag the correct D^0 flavour and therefore must be distinguished from the signal decays. Such events will not peak in Δm but rather have a smoothly increasing shape from the pion rest mass threshold as shown for simulation in Fig. 7.3(right). This type of background can be produced for both prompt and secondary D^0 candidates and thus there are two separate random soft pion components in the later fit stages where prompt and secondary contributions are distinguished.

Combinatorial background is also present in the Δm distribution. Its fraction with respect to the signal and random soft pion components is determined from fitting the $m(D^0)$ distribution. In practice, the Δm and $m(D^0)$ distributions are fitted together as two independent variables in a 2D fit. The same "root exponential" threshold PDF is used to fit combinatorial and random soft pion backgrounds, with independent parameters. This PDF is given by,

$$f(x) = \begin{cases} (\Delta m - A)^B \times \exp(C \cdot \Delta m) & \Delta m \geq A \\ 0 & \Delta m < A \end{cases}, \tag{7.4}$$

where A, B and C are referred to as the threshold, power and exponential decay parameters, respectively. The random soft pion PDF threshold parameter (A_{π_s}) is set to the pion PDG rest mass and the exponential decay parameter (C_{π_s}) is fixed to 0, such that the power parameter (B_{π_s}) is the only floating parameter. For combinatorial background the threshold parameter ($A_{comb.}$) is again fixed to the pion PDG rest mass, but the other two parameters ($B_{comb.}$,$C_{comb.}$) are allowed to float, giving two free parameters. For the $D^0 \to K^-\pi^+$ decay, the background is relatively small compared to the number of signal events. As a result, the combinatorial background component is simplified by fixing $C_{comb.}$ to be zero.

The correctly tagged signal component in Δm is made of three Gaussian functions, two of which share the same mean ($\mu_{\Delta m_1, sig.}$). Thus there are seven fit parameters for the signal component, including three widths ($\sigma_{\Delta m_i, sig.}$) and two relative fractions ($P(\Delta m_i|sig.)$).

The Δm projection of the 2D fit to $m(D^0)$ and Δm is shown in Fig. 7.2(right) for one K^+K^- sub-dataset. The fit parameters and values for this one sample are shown in Table 7.2.

7.4 Specific backgrounds

Specific backgrounds are those that appear in only one of the three final states of interest. There are two sources of specific backgrounds. The first source is "reflections", where true two-body D^0 decays are reconstructed under the wrong mass hypothesis due to particle misidentification. The second and largest source is "partially reconstructed decays", where two of the daughter particles of a multibody[1] charm decay have been identified as D^0 daughters and the other particles have not been reconstructed. This can occur with or without misidentification of the daughter particles. Backgrounds from real D^0 decays have a peaking component in Δm as well as a component from the association of a random soft pion. If one or several D^0 decay products were not reconstructed, the resulting peak in Δm is significantly wider than that for fully reconstructed decays. Backgrounds from particles other than a D^0 cannot peak in Δm and therefore

[1] A multibody decay refers to a decay with more than two daughter products.

only have a random soft pion component. Like combinatorial background from B decays, partially reconstructed charm decays are expected to be suppressed by requirements on the momentum direction of the reconstructed two-body decay and vertex quality requirements.

Backgrounds specific to $D^0 \to K^-\pi^+$

A potential background in $D^0 \to K^-\pi^+$ decays is the reflection of $D^0 \to K^+K^-$ decays, where one of the kaons is misidentified as a pion. The background appears in the low $m(D^0)$ sideband due to the loss of invariant mass. The factor 10 suppression from the relative branching fractions of these decays and the excellent particle identification efficiencies lead to this being a small background. Preliminary fits to the $K^-\pi^+$ data found that no significant K^+K^- reflection component was present. Similarly, a contribution of $D^0 \to \pi^+\pi^-$ decays may be expected in the high $m(D^0)$ sideband. However, this is suppressed by a factor of ~ 30 due to its branching fraction, and is further suppressed to a negligible level by particle identification requirements. A previous study of these backgrounds in [75], found that only the misreconstructed decay $D^0 \to \pi^-\pi^+\pi^0$ has a potentially significant contribution to the data. This is therefore the only specific background considered in the fit for $D^0 \to K^-\pi^+$. However, its fraction in preliminary fits was consistent with zero, and it was therefore fixed to zero for the final analysis fits.

Backgrounds specific to $D^0 \to K^+K^-$

The events selected under the K^+K^- hypothesis are most likely to be affected by backgrounds. This is because imposing the kaon mass hypothesis on a lighter particle can compensate for a loss in invariant mass when only partially reconstructing a decay. The favoured $D^0 \to K^-\pi^+$ decay lies outside the $m(D^0)$ range and can be ignored. A previous study of these backgrounds in [75], found that four partially reconstructed decays contribute significantly to the data. The first of these decays are, $D^0 \to K^-\pi^+\pi^0$ and $D^0 \to K^-\mu^+\nu_\mu$ where the neutral particle is not reconstructed and one of the other daughters is misidentified. The distribution of these events is found to be similar for both backgrounds so that only one component is needed in the fit. The other two partially reconstructed backgrounds are $D^+ \to K^-\pi^+\pi^+$ and $D_s^+ \to K^+K^-\pi^+$. The D^+ decay results in the same mass distributions as the random soft pion background and is absorbed into that component in the fit, while a separate D_s^+ component is included in

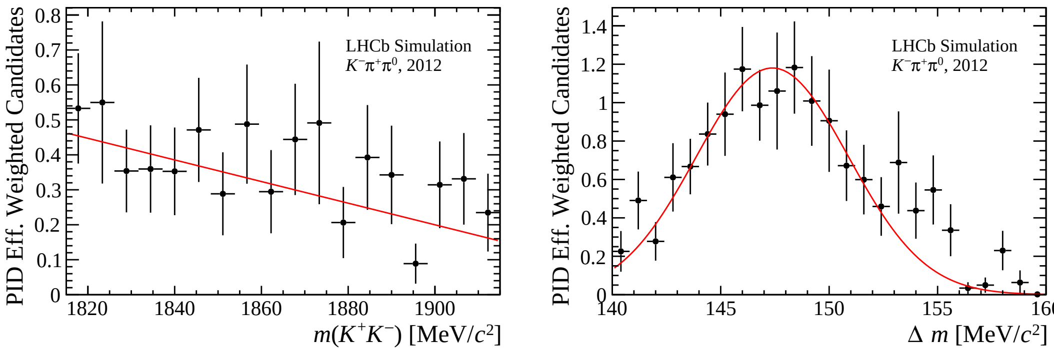


Figure 7.4: Fits to (left) $m(D^0)$ and (right) Δm of the $K^-\pi^+\pi^0$ decay from simulation, reconstructed as K^+K^- and PID efficiency weighted.

the fit. The sum of the $D^0 \to K^-\pi^+\pi^0$ and $D^0 \to K^-\mu^+\nu_\mu$ decays amount to approximately 3% of the total data, while the D_s^+ contributes approximately 1% of the events in the total data.

Backgrounds specific to $D^0 \to \pi^+\pi^-$

The only significant specific background to $D^0 \to \pi^+\pi^-$ is the reflection from $D^0 \to K^-\pi^+$ decays. Due to the loss in invariant mass these sit in the low-mass sideband of the D^0 mass distribution. Even though the $D^0 \to K^-\pi^+$ decay has a much larger branching fraction than the $D^0 \to \pi^+\pi^-$ decay, its contribution to the $D^0 \to \pi^+\pi^-$ mass peak is greatly reduced by the excellent particle identification of the experiment. Preliminary fits to the $\pi^+\pi^-$ data found that no significant $K^-\pi^+$ reflection component was present and hence it is neglected. The high-mass sideband of the D^0 mass distribution is free of specific backgrounds.

Specific background mass fit PDFs

To study the distributions of these backgrounds in $m(D^0)$ and Δm, simulated events were fitted after applying the stripping selection. Applying the full stripping selection returns too few candidates to obtain accurate fits, and consequently the PID requirements were not applied to these simulation samples. The PID efficiency for each background was calculated using the PIDCalib package [76], which contains reference samples of known particle types. The PID efficiency of the reference samples are binned in track η and momentum, p, using the PID selection requirements of the stripping selection as input. These efficiencies are used to determine the per-event PID efficiency of the simulation sample particles, given their η, momentum, particle type and using the two candidate daughters. The

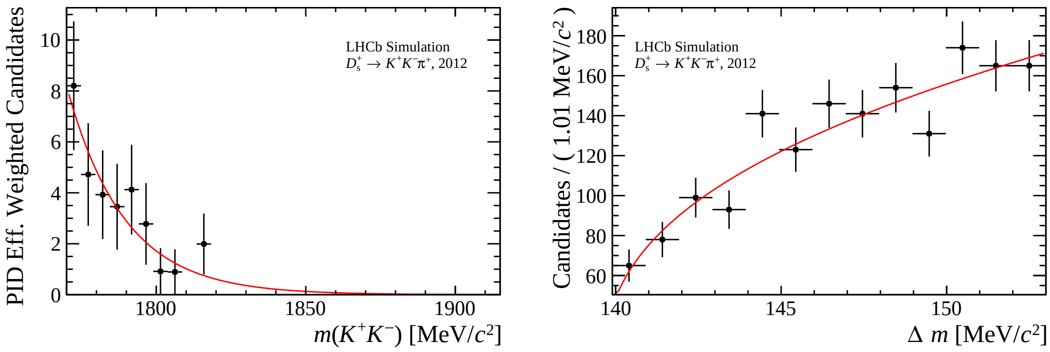


Figure 7.5: Fits to (left) $m(D^0)$ and (right) Δm of the D_s^+ decays from simulation, reconstructed as K^+K^-. The mass plot has the stripping requirements applied and is PID efficiency weighted while the Δm plot has no stripping requirements applied and is not reweighted.

D^0 mass and Δm of the specific background simulation samples are reweighted with the efficiencies before fitting.

Fits to the $D^0 \to K^-\pi^+\pi^0$ simulation can be seen in Fig. 7.4. The distribution has a linear shape in the signal mass region, as shown in the figure, and is fitted with a single gradient ($g_{K\pi\pi}$). The Δm distribution of this background is fitted with a single Gaussian PDF ($\mu_{\Delta m,K\pi\pi}$, $\sigma_{\Delta m,K\pi\pi}$). In the main fit a random soft pion component is also added, using fits to the well known random soft pion background shape and using a PDF given by,

$$f(x) = \begin{cases} \left(1 - \exp\{-\frac{x-x_0}{c}\}\right)\left(\frac{x}{x_0}\right)^a + b\left(\frac{x}{x_0} - 1\right) & x \geq x_0 \\ 0 & x < x_0 \end{cases}, \tag{7.5}$$

which is similar to the root exponential threshold PDF. This PDF has four parameters; $x_{0,(K\pi\pi)}$ sets the threshold, while $a_{(K\pi\pi)}$, $b_{(K\pi\pi)}$ and $c_{(K\pi\pi)}$ control the shape. It is not analytically integrable except for particular values of the floating parameters, thus a numerical integration is performed when normalising.

The D_s^+ mass, as shown in Fig. 7.5, was fitted with an exponential having one decay parameter ($e_{D_s^+}$). After PID efficiency weighting, there were too few weighted events in the Δm range considered for this analysis. The shape of the D_s^+ background in this variable was taken from a fit to the simulated sample without stripping requirements applied and using the same PDF as in Eq. (7.5).

In the main fit the shapes of the $K^-\pi^+\pi^0$ and D_s^+ backgrounds are fixed to those shown in the figures, whilst the fraction of their contributions to the data are allowed to float.

Table 7.2: Fit parameter results for the fit to $m(D^0)$ and Δm from one $D^0 \to K^+K^-$ sub-dataset with magnet polarity down and from the second run block. The symbols are explained in the text. The units of the parameters are shown where relevant. In the case where a formula is presented in the first column, the value in the last column is the X variable of the formula. For example, $\sigma_{m_2,sig.} = (1.745)\ \sigma_{m_1,sig.}$.

Parameter	Fixed?	Value
$P(sig.)$	N	(0.6227 ± 0.0012)
$P(\pi_s)$	N	(0.2664 ± 0.0012)
$P(K^-\pi^+\pi^0)$	N	(0.0316 ± 0.0035)
$P(D_s^+)$	N	(0.0093 ± 0.0011)
$P(m_1 \vert sig.)$	N	(0.791 ± 0.010)
$\mu_{m,sig.}$	N	$(1866.6240 \pm 0.0084)\,\mathrm{MeV}/c^2$
$\sigma_{m_1,sig.}$	N	$(6.188 \pm 0.028)\,\mathrm{MeV}/c^2$
$\sigma_{m_2,sig.} = X\ \sigma_{m_1,sig.}$	N	(1.745 ± 0.019)
$e_{comb.}$	N	$(-0.00486 \pm 0.00067)[\mathrm{MeV}/c^2]^{-1}$
$e_{D_s^+}$	Y	$-0.052\ [\mathrm{MeV}/c^2]^{-1}$
$g_{K\pi\pi}$	Y	-0.0001
$P(\Delta m_1 \vert sig.)$	N	(0.174 ± 0.024)
$P(\Delta m_2 \vert sig.)$	N	(0.576 ± 0.014)
$\mu_{\Delta m_1,sig.}$	N	$(145.4186 \pm 0.0017)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_1,sig.}$	N	$(0.365 \pm 0.015)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_2,sig.} = X\ \sigma_{\Delta m_1,sig.}$	N	(1.880 ± 0.040)
$\mu_{\Delta m_3,sig.}$	N	$(145.592 \pm 0.017)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_3,sig.} = X\ \sigma_{\Delta m_2,sig.}$	N	(1.923 ± 0.029)
A_{π_s}	Y	$139.57\,\mathrm{MeV}/c^2$
B_{π_s}	N	(0.3652 ± 0.0038)
C_{π_s}	Y	$0\ [\mathrm{MeV}/c^2]^{-1}$
$A_{comb.}$	Y	$139.57\,\mathrm{MeV}/c^2$
$B_{comb.}$	N	(0.756 ± 0.042)
$C_{comb.}$	N	$(116.4 \pm 7.2)[\mathrm{MeV}/c^2]^{-1}$
$\mu_{\Delta m,K\pi\pi}$	Y	$147.38\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m,K\pi\pi}$	Y	$3.52\,\mathrm{MeV}/c^2$
$x_{0,(D_s^+)}$	Y	$139.57\,\mathrm{MeV}/c^2$
$a_{(D_s^+)}$	Y	5.8970
$b_{(D_s^+)}$	Y	$0.01\,\mathrm{MeV}/c^2$
$c_{(D_s^+)}$	Y	$1.6252\,\mathrm{MeV}/c^2$
$x_{0,(K\pi\pi)}$	Y	$139.57\,\mathrm{MeV}/c^2$
$a_{(K\pi\pi)}$	Y	1.5266
$b_{(K\pi\pi)}$	Y	$0.01\,\mathrm{MeV}/c^2$
$c_{(K\pi\pi)}$	Y	$2.0453\,\mathrm{MeV}/c^2$

Chapter 8

Acceptance correction and decay-time fit

In this chapter, the second fit stage of the analysis is presented along with the procedure to account for the decay-time biases discussed previously in Section 5.4. Recall that the second fit stage is a 2D fit to the D^0 decay time and $\ln(\chi^2_{\text{IP}}(D^0))$. In Section 8.1 the secondary background events and the observable used to distinguish them in the fit are discussed in detail. The "swimming" algorithm for determining per-event decay-time biases is presented in Section 8.2, while additional acceptance effects are discussed in Section 8.3. The inclusion of these acceptance effects in the fit is discussed in Sections 8.4 and 8.5. Section 8.6 presents the PDF models used to describe the decay-time distributions of the classes and subclasses of the analysis, and Section 8.7 presents the PDF models in the $\ln(\chi^2_{\text{IP}}(D^0))$ observable. Finally, the construction of background components for which a PDF model is unknown is performed in Section 8.8.

8.1 Backgrounds in $\ln(\chi^2_{\text{IP}}(D^0))$

As introduced in Chapter 4, the data contains a background of secondary D^0 candidates that originate from the decay of a long-lived particle, predominantly B decays. The B in such a decay is not reconstructed and thus the candidate D^0 flight distance is overestimated, as was shown in Fig. 4.2. B meson decay times are typically $\sim 1.5\,\text{ps}$, which results in a significant increase in the reconstructed D^0 candidate decay time.

The B flight distance in secondary decays is not reconstructed because the D^0 flight distance is defined as the distance between the D^0 vertex and the

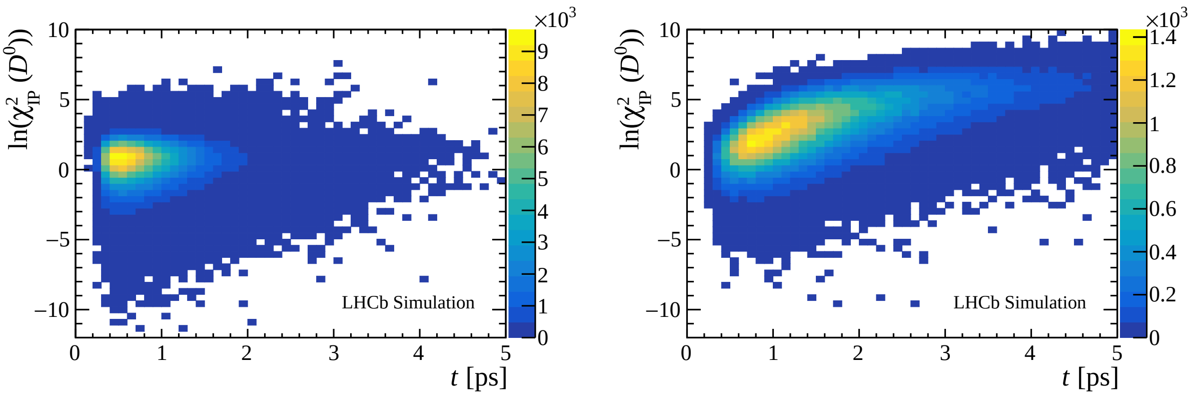


Figure 8.1: Distribution of decay time, t, against $\ln(\chi^2_{\rm IP}(D^0))$ for simulated events with magnet polarity down and 2012 conditions for (left) prompt signal and (right) secondary background. The decay-time dependence of the $\ln(\chi^2_{\rm IP}(D^0))$ of secondary decays is evident.

PV, instead of the distance between the D^0 vertex and the B vertex. If the secondary decay was fully reconstructed, the correct D^0 flight displacement could be determined. The analysis of a D sample arising from B decays provides an alternative method of performing this analysis. The B vertex can be determined by using another tagging particle such as a muon [35]. However, the statistics of such an analysis and thus the precision on the final result are significantly reduced. For an analysis of the prompt sample, as in this thesis, there will always be some secondary background that must be accounted for in the fit method, unless all secondary decays can be reconstructed. It is not possible to explicitly construct each of these decays as there are numerous channels, some of which contain neutrinos and neutral particles that pose a challenge for LHCb.

Secondary decays are indistinguishable from the prompt signal in most observables as they are real D^0 and real D^* decays. Due to the B lifetime and its associated flight distance, the $\chi^2_{\rm IP}$ of secondaries will be decay-time dependent. A larger B flight distance will result in a larger IP_{D^0} and a larger $\chi^2_{\rm IP}(D^0)$. The variable $\ln(\chi^2_{\rm IP}(D^0))$ is used in the fit procedure to distinguish prompt and secondary decays as it is easier to model than IP_{D^0} or $\chi^2_{\rm IP}(D^0)$, which both decrease exponentially and have no peaking structures. Fig. 8.1 shows the D^0 decay time against $\ln(\chi^2_{\rm IP}(D^0))$ for prompt signal and secondary background simulated events. The secondary background can be distinguished from the prompt signal in the fit from its decay-time dependence in $\ln(\chi^2_{\rm IP}(D^0))$.

8.2 Swimming algorithm

As outlined in Section 5.4, non-trivial decay-time acceptances are introduced by the selection procedure for candidate events. The method used in this analysis to measure the decay-time acceptances is the so-called "swimming" algorithm, which is a data-driven technique that was originally developed for the CDF experiment [77, 78] and was later studied and implemented at LHCb [79,80]. This section describes the current implementation of the swimming algorithm within the LHCb software.

The swimming algorithm works by artificially changing the decay-time of each real data event and recalculating the selection decision based on the new reconstruction of the event. This is done for a large set of discrete decay-time values to determine the decay-times at which the event would have been accepted by the selection criteria and when it would not have been selected. In this way, the acceptance function for each event is the sum of "top-hat" or "step" functions. The decay times at which the event enters or leaves one of these top hat functions are called turning points (TPs). The acceptance function can be represented as a vector of turn on and subsequent turn off points. The decay-time of each event is varied by changing the distance between the PV and the SV. Figs. 8.2 and 8.3 show schematics for this procedure using IP selection requirements as an example.

For each event there are two main decay-time biasing parts of the selection to consider; the trigger selection and the stripping selection. In addition, it is necessary to account for the acceptance due to the finite length of the VELO, as well as any potential dependence of the track finding efficiency on the D^0 decay time.

Each event is "swum" independently for each source of acceptance to determine the per-event and per-source acceptance functions. In order to obtain the total acceptance function for each event the various sources are merged. The final acceptance function for a single event has a value of one (or "on") if all of the individual acceptance functions have a value of one, and has a value of zero (or "off") if any of the individual functions have a value of zero, at a certain decay time.

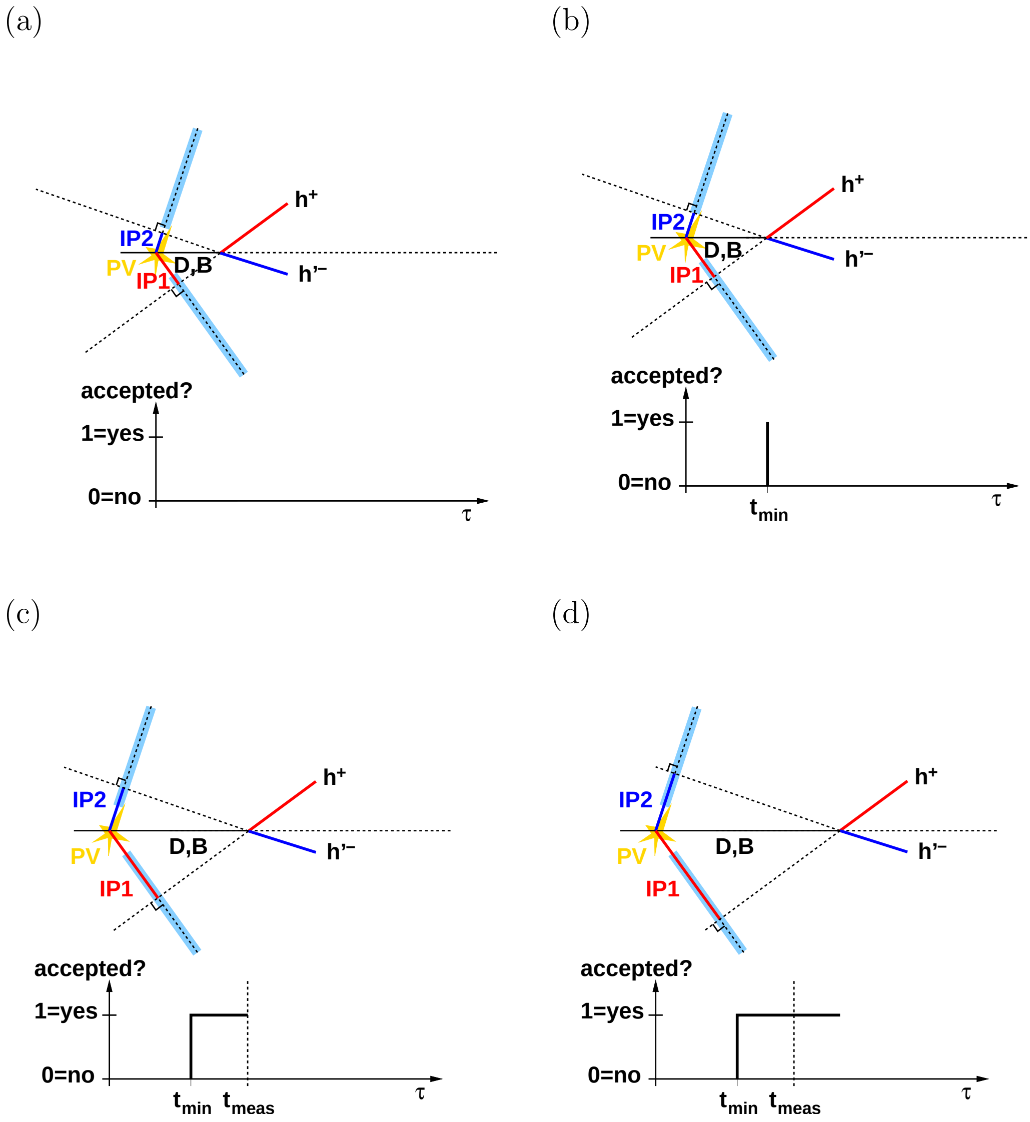


Figure 8.2: Schematic of the swimming method for a two-body hadronic decay and with IP selection requirements as an example. Shown below each schematic is a graph of the decay-time acceptance function for this event. The shaded, light blue regions show the bands for accepting the IP value of the daughter tracks. As discussed in Chapter 5, no selection is made on IP, however χ^2_{IP} selections are made that bias the lifetime in a similar manner. In (a) the h^- daughter track fails the IP requirement and the acceptance is turned off. The distance between the PV and SV is increased in (b) until the both daughter tracks pass the IP requirement and the acceptance turns on at t_{min}. In (c) and (d) the decay-time is further increased and the acceptance remains on. The decay-time measured by the experiment, t_{meas} should lie within the acceptance intervals as in (d).

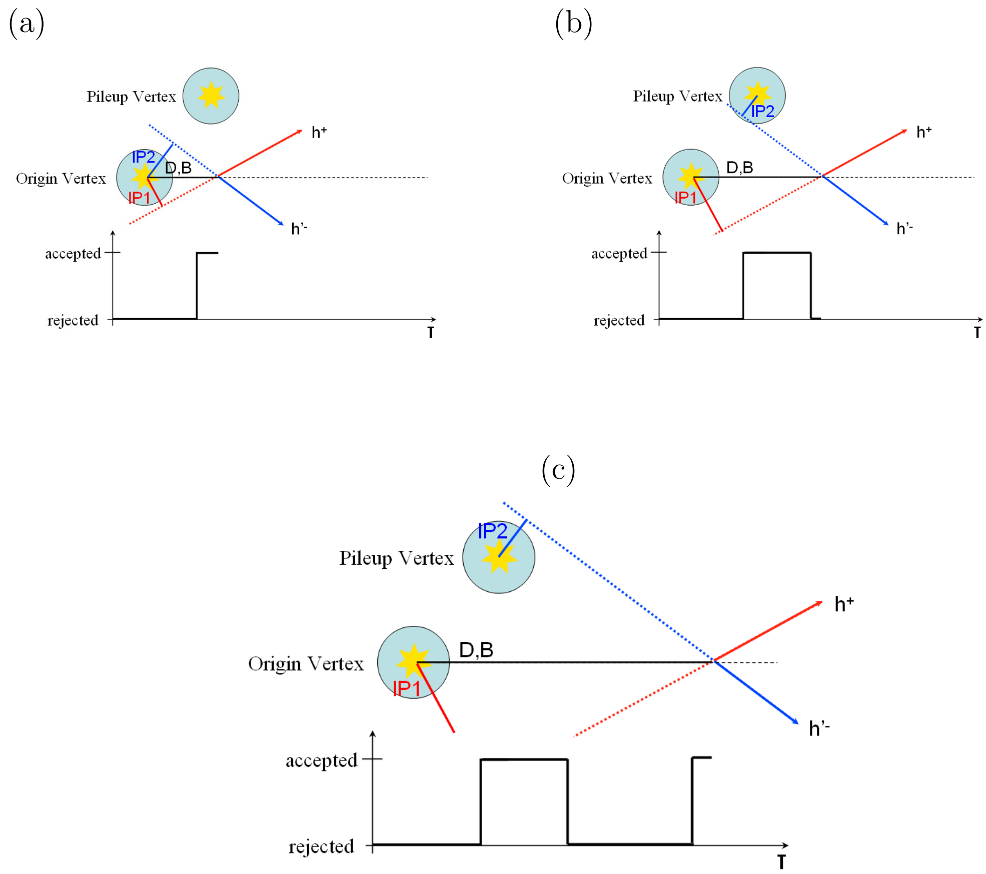


Figure 8.3: Schematic of the swimming method for a two-body hadronic decay and with IP selection requirements as an example where there are two PVs in the event. Shown below each schematic is a graph of the decay-time acceptance function for this event. The circles about each PV show the minimum IP value that the daughter tracks must have for the event to be selected. The IP values are large enough for the event to be accepted in (a) and thus the acceptance is turned on. In (b) the event is not accepted as the minimum IP for the h^- daughter is now associated to the second PV and is smaller than the accepted value. In (c) the acceptance turns on again as the IP w.r.t. the second PV is larger than the selection requirement.

8.2.1 Swimming the trigger

Section 5.1 described the trigger selection used for this analysis. Swimming the trigger involves rerunning this trigger algorithm for each event while varying the D^0 decay time and measuring whether the event passed or failed the trigger selection at each value of the decay-time. The decay-time is varied by moving the primary vertex along the D^0 momentum direction. The D^0 decay vertex is not varied as this would require a full reconstruction of the daughter particles. All primary vertices in an event are moved coherently in the direction of the D^0 momentum, and the PV associated with the D^0 is recalculated at each swimming step, thus taking into account acceptance effects due to PV misassociation.

The key to this method is the ability to rerun the trigger in precisely the same configuration as during data taking. This is made possible by the implementation of the trigger as a software framework, instead of a hardware device. In order to reproduce the trigger decisions as performed during data taking, it is sufficient to choose the same version of Moore, the same trigger configuration key (TCK), which encodes the selection requirements in each trigger line, and the correct entries from a database of the detector conditions and alignment.

Each decay time, t_i, for a single event, has a one-to-one correspondence with the positions of the PVs in the event. Given infinite computing resources one could evaluate the trigger decision at every possible decay time of the D^0 candidate and obtain the acceptance function to some arbitrary precision. In practice this is not possible because the number of times that the trigger would have to be evaluated for each event becomes larger as the sampling granularity increases. Instead a search procedure is carried out to increase precision and decrease computation time. First, a rough search is performed for the turning points, with a granularity of 4 mm between ± 200 mm of the original primary vertex position and a granularity of 40 mm between ± 600 and ± 200 mm. A turning point is defined by the change in the TOS trigger decision between two consecutive swimming steps. The granularities and swimming distances are chosen because a D^0 with a momentum of 200 GeV/c flies approximately 20 mm in the LHCb lab frame, and it is necessary to know the acceptance function to at least ± 10 D^0 lifetimes. Beyond this region (200 to 600 mm), the swimming is performed with a coarser granularity to allow upper lifetime acceptances to be tested. The granularity of 4 mm means that top hat acceptances narrower than 4 mm may be missed by this procedure. Such narrow top hat functions are indicative of overlapping primary vertices as the D^0 must first fly far enough away from one

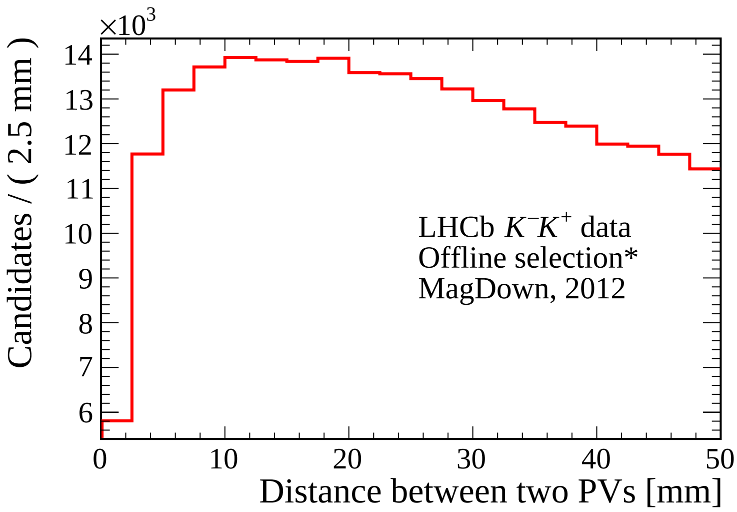


Figure 8.4: Distribution of the distance between the two PVs in events with only two PVs.

primary vertex to be accepted, only to be rejected for being too close to another primary vertex shortly afterwards. Given the size of the interaction region of $\pm 30\,\mathrm{mm}$ and the average primary interaction multiplicity of two, it was felt that the primary vertices were well enough separated to make this effect negligible (see Fig. 8.4).

In the next step of the search procedure, each turning point is refined by searching in a geometrically decreasing interval around itself; first in steps of 1 mm in the region of $\pm 4\,\mathrm{mm}$ around the turning point, then 250 µm and so on. Four refinement steps are used, giving a final turning point precision of $\frac{15.625\ \mu\mathrm{m}}{\sqrt{12}}$.

8.2.2 Swimming the stripping

The swimming of the stripping selection is done in the exact same way as the trigger swimming, with the same step granularities and number of iterations. Again, the stripping is a software framework and can thus be run multiple times with identical settings. To uniquely specify a stripping configuration, only the DaVinci version and correct detector conditions are required.

8.3 Further acceptance corrections

8.3.1 Accounting for the VELO acceptance

The finite length of the VELO can impose an upper lifetime acceptance on very long-lived D^0 candidates whose daughters are unlikely to traverse the required number of nine VELO stations at HLT1. The swimming algorithm is not sensitive to this effect as the decay time is varied by moving the PV instead of the D^0 decay vertex, and thus the daughter hits are not reconstructed. The VELO acceptance is taken into account by analytically computing the number of stations that a track passes through, based on the position of the D^0 decay vertex and the daughter track directions. This calculation includes subtleties in the VELO geometry such as tracks that cross the overlap between the left-hand and right-hand halves. As for the swimming algorithm, a series of TPs is produced that defines the acceptance intervals for each event due to these effects.

8.3.2 Swimming the radial flight distance requirement

As described in Section 5.4, the additional offline selection that biases the decay-time is the requirement that the D^0 radial flight distance is less than 4 mm. The effect of this selection on the decay time can be determined analytically from the D^0 kinematics. The flight distance vector is,

$$\mathbf{D} := \mathbf{V}_{D^0} - \mathbf{V}_{\mathrm{PV}}, \tag{8.1}$$

where $\mathbf{V}_{D^0}$ is the position of the D^0 decay vertex and $\mathbf{V}_{\mathrm{PV}}$ is the PV position. Defining $\mathbf{p}_{\mathrm{unit}}$ as the unit D^0 momentum vector and increasing the D^0 flight distance by k units, the radial flight distance squared is given by,

$$r^2 = (D_x + kp_{\mathrm{unit},x})^2 + (D_y + kp_{\mathrm{unit},y})^2. \tag{8.2}$$

As this is parabolic in k it can be solved for k in the usual way for a given r, where the larger of the two values of k represents the upper limit of the acceptance. The lower solution for k gives the lower limit of the acceptance, a negative decay-time that is excluded by the trigger requirements. Having obtained k, the maximum decay time allowed by the radial flight distance requirement is,

$$t_{\mathrm{max}} = t_{\mathrm{meas}} \frac{|\mathbf{D} + k\mathbf{p}_{\mathrm{unit}}|}{|\mathbf{D}|}, \tag{8.3}$$

where t_{meas} is the measured decay time of the candidate at its original flight distance.

8.3.3 Accounting for the tracking efficiency acceptance

In addition to the above effects the VELO track reconstruction creates a decay-time acceptance effect that is unconnected to the VELO length. The track finding algorithm prefers tracks that originate close to the beamline. Long-lived D^0 mesons are more likely to travel further from the beam, leading to a potential upper lifetime acceptance limit that is not accounted for by any of the previous swimming steps.

This effect was studied in [81], using $B^+ \to J/\psi K^+$ decays where the section of the kaon track downstream of the magnet is reconstructed, and subsequently the VELO part of the track is searched for as a function of the track distance of closest approach to the beamline, z_{doca}. The efficiency is parameterised as,

$$\epsilon(z_{\text{doca}}) = \alpha + \beta z_{\text{doca}}^2, \tag{8.4}$$

where α and β are taken from an appendix of the reference in which they are calculated in bins of the track transverse momentum and event occupancy.

In order to interpret these results as an acceptance effect on the D^0 decay time, a link must be made between the track z_{doca} and the decay time. This link is dependent on the kinematics and geometry of each event. The dependence of z_{doca} on the decay time is shown in Fig. 8.5 for a single event. The dependence is linear but the slope and intercept will vary for each event. Thus the efficiency in terms of decay time for a single track, i, is given by,

$$\epsilon_i(t) = \alpha + \beta(\gamma_i + \delta_i t)^2, \tag{8.5}$$

where γ_i and δ_i are the intercept and slope as in Fig. 8.5.

All four constants are known or can be computed for each daughter track in each event prior to the fit. The application of this in the analysis is discussed in Section 8.6.

8.4 Time-fit PDF

The total fit PDF factorised into the two fit parts is given in Eq. (7.2). Decay-time acceptance is included as a set A of the TPs returned from the swimming and other acceptance determinations. Only candidates with two turning points (a single top-hat acceptance function) are used in the analysis. This greatly simplifies the inclusion of the acceptance in the total fit PDF. The selection requirement on the D^0 decay radius means that the number of events with more than two TPs

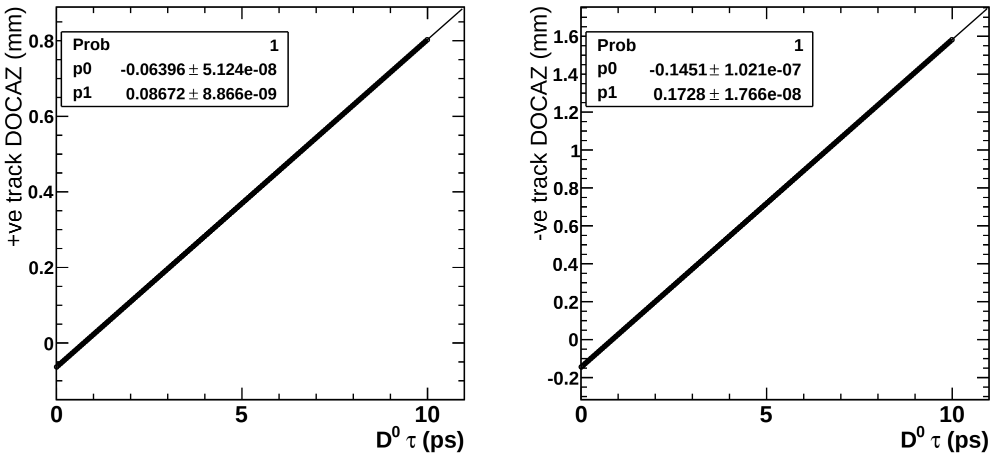


Figure 8.5: Dependence of track z_{doca} on D^0 decay time for one event, showing (left) the positively charged track and (right) the negatively charged track.

is small. To reduce correlation between the two turning points, the observable $\mathrm{TP}_{\mathrm{diff}} := \mathrm{TP}_2 - \mathrm{TP}_1$, is defined, such that $A \approx \{\mathrm{TP}_1, \mathrm{TP}_{\mathrm{diff}}\}$. These acceptance observables only appear in the decay-time PDFs, as acceptance effects on the mass and $\ln(\chi^2_{\mathrm{IP}})$ PDFs are not important to this measurement.

Including acceptance effects, the PDF of the second fit stage is given by,

$$\begin{aligned}\sum_{class} P(class|m,\Delta m) f_{class}(\ln(\chi^2_{\mathrm{IP}}), t, A) = \\ P(sig.|m,\Delta m) \sum_{subcl.} P(subcl.|sig.) f_{sig.,subcl.}(\ln(\chi^2_{\mathrm{IP}}), t|A) f_{sig.,subcl.}(A) \\ + P(\pi_s|m,\Delta m) \sum_{subcl.} P(subcl.|\pi_s) f_{\pi_s,subcl.}(\ln(\chi^2_{\mathrm{IP}}), t|A) f_{\pi_s,subcl.}(A) \\ + \sum_{kernel} P(kernel|m,\Delta m) f_{kernel}(\ln(\chi^2_{\mathrm{IP}}), t|A) f_{kernel}(A),\end{aligned} \tag{8.6}$$

where *kernel* refers to the combinatorial or misreconstructed backgrounds with no distinction between prompt and secondary decays. Recall that the relative fractions $P(class|m,\Delta m)$ are known from the first fit stage. The new relative fractions, $P(subcl.|class)$ describe the fraction of prompt and secondary decays within a class. They are floating parameters in the second fit stage, for which the $\ln(\chi^2_{\mathrm{IP}})$ distribution provides discrimination. The PDFs $f_{class,subcl.}(A)$, describe the distribution of TP_1 and $\mathrm{TP}_{\mathrm{diff}}$ for each subclass and their determination is described in Section 8.5. Each of the functions, $f_{class,subcl.}(\ln(\chi^2_{\mathrm{IP}}), t|A)$, can be further split into the decay time and $\ln(\chi^2_{\mathrm{IP}})$ components as follows,

$$f_{class,subcl.}(\ln(\chi^2_{\mathrm{IP}}), t|A) = f_{class,subcl.}(t|A) f_{class,subcl.}(\ln(\chi^2_{\mathrm{IP}})|t), \tag{8.7}$$

where acceptance effects on $\ln(\chi^2_{\mathrm{IP}})$ have been ignored and where $f_{class,subcl.}(t|A)$ is the decay-time PDF given the acceptance, discussed in Section 8.6. The last line in Eq. (8.6) for the kernel PDFs can similarly be factorised into a decay-time PDF and a decay-time dependent $\ln(\chi^2_{\mathrm{IP}})$ PDF. The turning point and decay-time PDFs of the kernel classes are not parametric and their acquisition is detailed in Section 8.8.

8.5 Turning point PDFs

The procedures for determining the PDFs of TP_1 and $\mathrm{TP}_{\mathrm{diff}}$ for the various classes and subclasses are presented here. These are the PDFs, $f_{class,subcl.}(A)$ in Eq. (8.6). As TP_1 and $\mathrm{TP}_{\mathrm{diff}}$ are correlated the PDF factorises as,

$$f_{class,subcl.}(\mathrm{TP}_1, \mathrm{TP}_{\mathrm{diff}}) = f_{class,subcl.}(\mathrm{TP}_1) f_{class,subcl.}(\mathrm{TP}_{\mathrm{diff}}|\mathrm{TP}_1). \tag{8.8}$$

If these PDFs were the same for all classes and subclasses they would factor out from Eq. (8.6), resulting in a constant multiplicative term. This term could then be neglected as it would have no effect on the position of the likelihood function maximum. However, the acceptance effects depend on the candidate kinematics and the kinematics vary between classes. Thus, in the fit it is necessary to include different PDFs describing TP_1 and $\mathrm{TP}_{\mathrm{diff}}$ for each class and subclass.

The procedure for determining these turning point PDFs for the signal and π_s classes, and their associated subclasses, is detailed in this section. The determination of the corresponding PDFs for the classes using kernel density estimates is introduced along with other details of these classes in Section 8.8. The PDFs resulting from these procedures are used in the final fit to the D^0 decay time and to $\ln(\chi^2_{\mathrm{IP}}(D^0))$. As a consequence, the iterative procedures described are run before the final analysis fit. They require information gained from the D^0 mass and Δm fit of Chapter 7, and thus are run after that step of the analysis.

In the case of signal and random π_s classes, the TP distributions of prompt and secondary decay components must be unfolded. This cannot be done using information from the mass fits only, as prompt and secondary decay candidates have identical distributions in $m(D^0)$ and Δm. The probability of a candidate

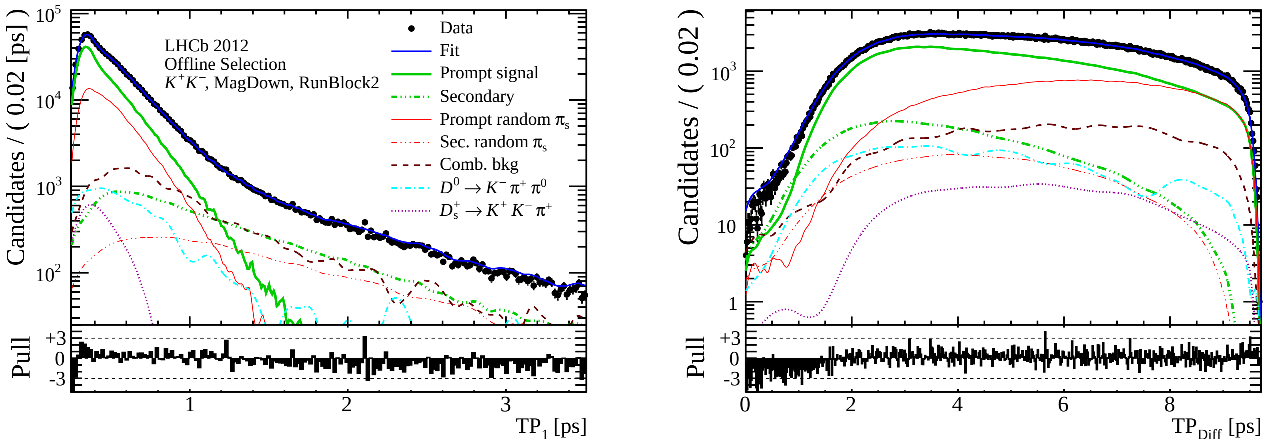


Figure 8.6: Comparisons of the TP PDF estimates summed over all classes and subclass with the TP distributions from data for (left) TP_1 and (right) TP_{diff}. Both estimated distributions show a good description of the data, though there is a small discrepancy between the estimate and the data at low decay times and small values of TP_{diff}. The determination of these PDFs is described in Section 8.5 for the signal and π_s classes and in Section 8.8 for the kernel classes.

belonging to a certain subclass (prompt or secondary) given its class (signal or random π_s) and its measured values, is given by,

$$P(subcl.|class, m, \Delta m, \ln(\chi^2_{\text{IP}}), t, \text{TP}_1) = \frac{P(subcl.|class) f_{class}(m, \Delta m) f_{class,subcl.}(\ln(\chi^2_{\text{IP}}), t|\text{TP}_1) f_{class,subcl.}(\text{TP}_1)}{\sum_{subcl.} P(subcl.|class) f_{class}(m, \Delta m) f_{class,subcl.}(\ln(\chi^2_{\text{IP}}), t|\text{TP}_1) f_{class,subcl.}(\text{TP}_1)}, \tag{8.9}$$

where lemma 2 in Section 6.2 is used. The mass PDFs for all classes are known from the first fit stage. An iterative unfolding procedure is used to determine the first turning point PDF from Eq. (8.9). Initial estimates of the first turning point distributions $f_{class,subcl.}(\text{TP}_1)$, are made from the data distribution of TP_1, sWeighted using the mass fit results. Kernel density estimation is applied to this distribution to produce a continuous function. To distinguish between the subclasses, an initial guess of $f_{class,prom.}(\text{TP}_1)$ is made from the data with $\ln(\chi^2_{\text{IP}}) < 1$, while the corresponding secondary decay PDF, $f_{class,sec.}(\text{TP}_1)$, is made with all candidates. As these probabilities are calculated prior to the decay-time fit, an initial estimate of the decay time and $\ln(\chi^2_{\text{IP}})$ PDF parameters, and the relative fractions must be made. Provided these are reasonably close (within 5σ) to the final fitted values this has negligible effect on the calculation of the probabilities. To minimise potential biases from the choice of starting parameters, a first fit is run with rough estimates of the fit parameters, and its output used to update the values of the parameters. This is then used as the default starting point for all fits. The initial estimate PDFs are used to calculate the first estimate of the probabilities, $P_1(subcl.|\dots)$. The turning point distributions are then remade using $P_1(subcl.|\dots)$ as the weights, instead of the sWeights, and without requirements on $\ln(\chi^2_{\mathrm{IP}})$. This is repeated until the difference between subsequent TP_1 PDFs is sufficiently small. The same procedure is used to determine the $\mathrm{TP}_{\mathrm{diff}}$ distribution, with suitable modifications to Eq. (8.9).

The accuracy of this method is confirmed by comparing the data to the sum of the obtained PDFs, as shown in Fig. 8.6 for K^+K^- data. The overall description of the data is good, and the differences in the PDFs between classes are apparent. In particular, secondary decay candidates have considerably larger values of TP_1 compared to prompt decay candidates. This is because D^0 mesons from B decays do not come from the PV and have larger direction angles at lower decay times. The DIRA selection requirement used in the stripping selection (See Table 5.3) is thus much more efficient for prompt decays than for secondary decays at low decay times.

8.6 Decay-time PDFs, $f_{class,subcl.}(t|A)$

Putting together the relevant terms described in the previous sections one can evaluate the PDFs for the decay time and $\ln(\chi^2_{\mathrm{IP}})$ fit. The decay-time PDF from Eq. (8.7) for a single event can be written as,

$$f_{class,subcl.}(t|\mathrm{TP}_1, \mathrm{TP}_{\mathrm{diff}}) = \frac{f_{class,subcl.}(t)\Theta\left(t - \mathrm{TP}_1\right)\Theta\left(\mathrm{TP}_1 + \mathrm{TP}_{\mathrm{diff}} - t\right)}{\int_{\mathrm{TP}_1}^{\mathrm{TP}_1+\mathrm{TP}_{\mathrm{diff}}} f_{class,subcl.}(t)\,\mathrm{d}t}, \quad (8.10)$$

where $\Theta(t)$ is the Heaviside step function and $f_{class,subcl.}(t)$ is the "unbiased" decay-time PDF. The biased decay-time PDF for a single event is the unbiased PDF restricted to, and normalised in, the acceptance interval, as in Fig. 8.7.

For the signal prompt class,

$$f_{sig.,prom.}(t) = \frac{1}{\tau}e^{-t'\tau} \otimes R(t', t)\ \ \epsilon_i(t), \quad (8.11)$$

where R is a detector resolution function and $\epsilon_i(t)$ is the track reconstruction efficiency for the prompt signal decays, first introduced in Section 8.3.3.

The detector resolution is accounted for by convolving the unbiased lifetime PDFs with a single Gaussian of width σ,

$$R(t', t) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t'-t)^2}{2\sigma^2}}. \quad (8.12)$$

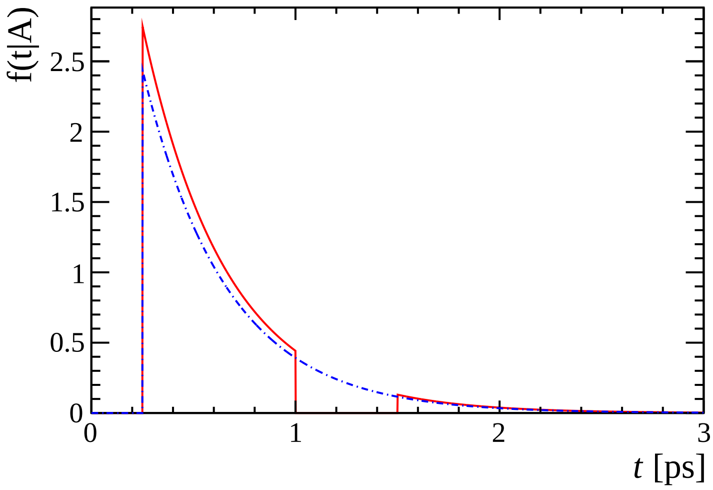


Figure 8.7: Example distribution of $f(t|A)$ for one event. The solid line shows the normalised distribution for which A contains three TPs at 0.25, 1, and 1.5 ps and a lifetime of 410.1 fs. The dashed line shows the normalised distribution of a pure exponential without acceptance effects and with the same lifetime.

This assumption is sufficient given the good resolution of LHCb. The σ value of 50 fs used in the fits is motivated by the averaged values measured with a triple Gaussian resolution model in three B decays, which lie between 43 fs and 51 fs [82]. The effect of the detector resolution on A_Γ is explored in Section 9.4.7 by varying the value of σ and by introducing a different resolution model. No detector resolution is applied to the secondary background component.

The track reconstruction efficiency as a function of t for a single track is given by Eq. (8.5). For a pair of tracks it is the product of the efficiency for each track (denoted by subscripts 1 and 2) leading to a fourth order polynomial in decay-time,

$$\begin{aligned}\epsilon_i(t) &= (\alpha + \beta z_{1,\mathrm{doca}}^2)(\alpha + \beta z_{2,\mathrm{doca}}^2) \\ &= (\alpha + \beta(\gamma_{1,i} + \delta_{1,i} t)^2)(\alpha + \beta(\gamma_{2,i} + \delta_{2,i} t)^2).\end{aligned} \tag{8.13}$$

This efficiency correction is not applied to the background PDFs.

The unbiased secondary background PDF is the sum of the convolution of two exponentials and a further single exponential,

$$f_{sec.}(t) = \frac{P(sec.1|sec.)\Big(e^{-t/\tau_{2,sec.}} - e^{-t/\tau_{1,sec.}}\Big) + \Big(1 - P(sec.1|sec.)\Big)e^{-t/\tau_{3,sec.}}}{P(sec.1|sec.)\Big(\tau_{2,sec.} - \tau_{1,sec.}\Big) + \Big(1 - P(sec.1|sec.)\Big)\tau_{3,sec.}}, \tag{8.14}$$

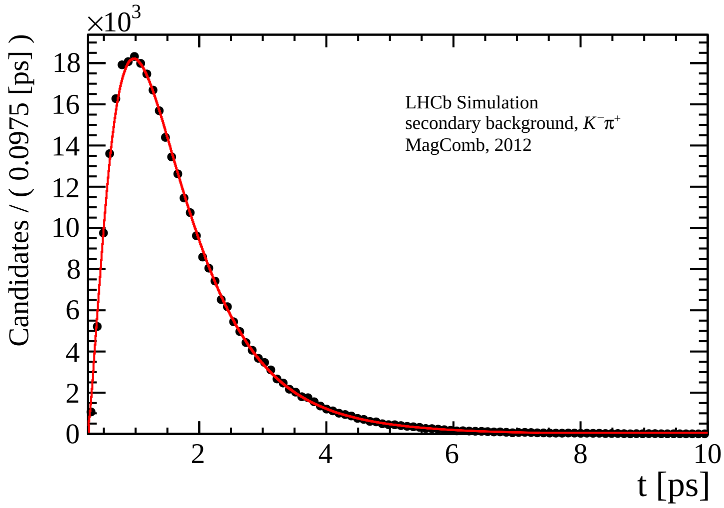


Figure 8.8: Fit to the D^0 decay time of a simulated secondary background sample decaying to the $K^-\pi^+$ final state. The full offline selection was applied to the sample. A convolution of two exponentials plus a third exponential is used as the fit function.

where $P(sec.1|sec.)$ is the fraction of the convolution with respect to the single exponential. The convolution of two exponentials gives this difference, which is motivated by the fact that there are two component parts to the decay time of these events, one from the B and one from the D^0. The parameter $\tau_{1,sec.}$ is fixed to 0.33 ps from a fit to a sample of simulated secondary decays, shown in Fig. 8.8. The parameter $\tau_{2,sec.}$ is constrained to be larger than $\tau_{1,sec.}$, as seen in the simulation fits. The addition of the third exponential improves the quality of the fits to simulation. The lifetimes of the secondary background PDFs returned from the fit are smaller than expected considering that B decays are relatively long-lived. This is due to the projection of the b-quark hadron flight displacement onto the D^0 flight displacement, which results in a smaller effective flight distance than would be expected.

The unbiased decay-time PDF of the prompt soft pion background follows exactly the same prescription as that of the prompt signal. The track reconstruction efficiency, $\epsilon_i(t)$, is assumed to be the same for both, although the fitted lifetime parameter, τ_{π_s}, is allowed to be different. The secondary soft pion background has the same unbiased decay-time PDF as the secondary signal sample with the same parameter values.

The $\ln(\chi^2_{\text{IP}})$ and D^0 decay-time observables are fit together and thus the two relative fractions for the subclass PDFs, $P(prom.|sig.)$ and $P(prom.|\pi_s)$, are

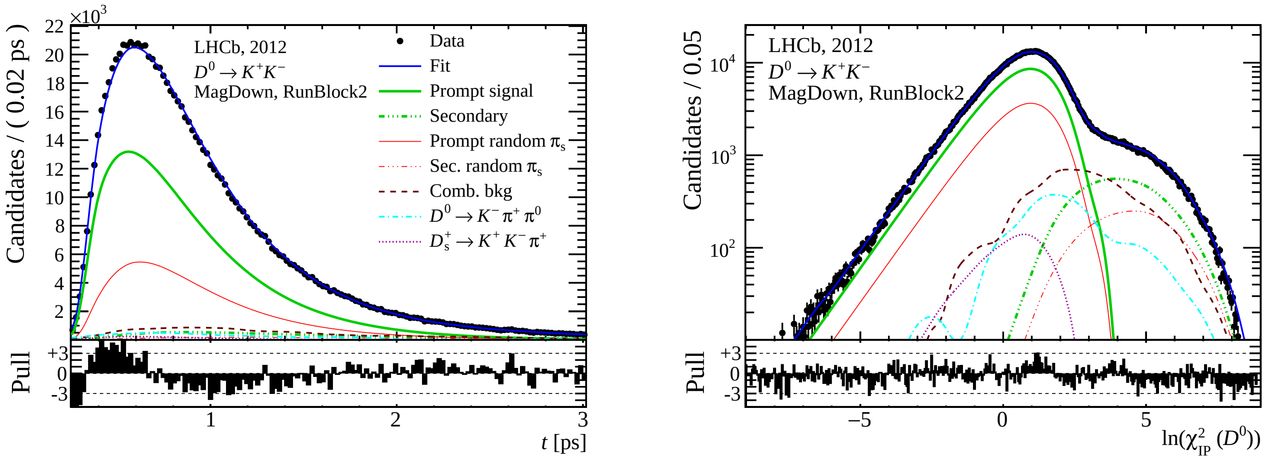


Figure 8.9: Projections of the fit to (left) D^0 decay-time and (right) $\ln(\chi^2_{\rm IP})$ for the $D^0 \to K^+K^-$, magnet polarity down sub-dataset in the second run block.

shared. By construction these parameters also control the fraction of secondary events in each class.

The D^0 decay-time projection of the fit to decay-time and $\ln(\chi^2_{\rm IP})$ for one sub-dataset is shown in Fig. 8.9(left). Table 8.1 shows the final values of all fit parameters for this one sub-dataset. Detailed results for all final states can be found in Appendix B. Discrepancies in the agreement between the fit and the data can be seen at low decay-times in the figure. This is discussed further in Section 9.4.4.

8.7 $\ln(\chi^2_{\rm IP})$ PDFs, $f_{class,subcl.}(\ln(\chi^2_{\rm IP})|t)$

The $\ln(\chi^2_{\rm IP})$ distributions in the fit are modelled with a "smeared $\ln(\chi^2_{\rm IP})$" PDF, given by,

$$f(x) = \begin{cases} \exp\left\{\alpha_L x - \exp\left\{\alpha_L(x-\mu_\chi)\right\}\right\} & x \leq \mu_\chi \\ \exp\left\{\alpha_L \mu_\chi + \alpha_R(x-\mu_\chi) - \exp\left\{\alpha_R(x-\mu_\chi)\right\}\right\} & x > \mu_\chi \end{cases}, \qquad (8.15)$$

where μ_χ is the most probable value and $\alpha_{L,R}$ are left and right-hand smearing parameters. The form of the PDF presented here has not been normalised but an analytic integration is performed in the analysis framework. The prompt signal $\ln(\chi^2_{\rm IP})$ component is described by two of such PDFs. A small decay-time dependence is seen in the simulated data in the shape of the prompt signal component, which is accounted for in the fit. Both mean parameters have a linear dependence on time, with different intercept parameters, but a shared slope,

$$\mu_{\chi,i,prom.}(t) = S_{\mu,prom.}x + I_{\mu,i,prom.}, \tag{8.16}$$

where all three parameters are allowed to float in the fit. The left-hand side smearing parameter of the first PDF is linear in time with a floating intercept and slope ($S_{\alpha_L,prom.}$, $I_{\alpha_L,prom.}$). The left-hand side smearing parameter of the second PDF is fixed ($\alpha_{L,2,prom.}$). The two right-hand side smearing parameters are constant in time, but their values are allowed to float ($\alpha_{R,i,prom.}$). Including a relative fraction for the first PDF, defined as $P(\chi 1|prom.)$, gives eight free parameters. The starting values and decay-time dependencies for these PDFs are motivated by studying the variation of the PDF parameters in bins of decay time with simulated data, as detailed in Appendix A.1. For the $\pi^+\pi^-$ final state, the prompt $\ln(\chi^2_{\text{IP}})$ PDF is a single smeared $\ln(\chi^2_{\text{IP}})$ PDF instead of the sum of two, as fewer free parameters are needed to fit this relatively small dataset.

A single smeared $\ln(\chi^2_{\text{IP}})$ PDF is used to represent secondary background. As mentioned in Section 8.1, the evident decay-time dependence of secondary background events is used to distinguish them from prompt signal events. The mean has a dependence on decay-time that is described by the sigmoid function,

$$\mu_{\chi,sec.}(t) = g_0 + \frac{g_2}{1 + e^{-g_1 t}}, \tag{8.17}$$

where all parameters are allowed to float. The left and right-hand side smearing parameters have a time evolution given by exponential functions,

$$\alpha_{L,sec.} = p_{L,0} + p_{L,1}t + e^{p_{L,2}t}, \tag{8.18}$$

$$\alpha_{R,sec.} = p_{R,0} + p_{R,1}t + e^{p_{R,2}t}, \tag{8.19}$$

The $p_{R,2}$ parameter is fixed for all final states, and thus there are eight free parameters from this component. The form of these time dependencies and their starting values come from fits to simulated secondary events, as described in Appendix A.2.

The prompt and secondary random soft pion components have the same PDFs as their associated correctly tagged components with the exact same parameter values.

The $\ln(\chi^2_{\text{IP}})$ decay-time projection of the fit to decay-time and $\ln(\chi^2_{\text{IP}})$ for one sub-dataset is shown in Fig. 8.9(right). Table 8.1 shows the final values of all fit parameters for this one sub-dataset.

8.8 Kernel density PDFs

The determinations of the kernel class background PDFs describing, the turning points, the $\ln(\chi^2_{\mathrm{IP}})$ distribution, and the unbiased decay time, are significantly different to those used for the other classes and subclasses of the analysis. Described here are the determinations of the same PDFs as in previous sections, but for the kernel classes.

The shapes of the combinatorial and misreconstructed backgrounds in terms of decay time and $\ln(\chi^2_{\mathrm{IP}})$ are unknown. Their PDFs in the fit are therefore kernelised sPlots, as introduced in Section 6.4. The sWeights of the different background classes in decay time and $\ln(\chi^2_{\mathrm{IP}})$ are extracted from the first fit stage to $m(D^0)$ and Δm. The biased decay time and $\ln(\chi^2_{\mathrm{IP}}(D^0))$ PDFs are obtained by first applying the sWeights for a certain class to the full dataset, giving the disjunct distributions of decay time and $\ln(\chi^2_{\mathrm{IP}}(D^0))$ for this class. Next, Gaussian kernels are applied to each event to form a continuous distribution of the biased decay time and $\ln(\chi^2_{\mathrm{IP}})$ PDFs for this class. The PDFs of $f(\mathrm{TP}_1)$ and $f(\mathrm{TP}_{\mathrm{diff}})$ are also determined in this way.

The unbiased decay-time PDF is extracted from the biased PDF using an iterative procedure. The biased and unbiased decay-time PDFs are related by,

$$f_{kernel}(t, A) = f_{kernel}(A|t) f_{kernel}(t). \tag{8.20}$$

For the background kernel classes in question, the acceptance PDF given the decay time is defined as,

$$f_{kernel}(A|t) = {}_s\mathcal{P}_{kernel} \frac{\Theta\left(t - TP_1\right)\Theta\left(TP_1 + TP_{\mathrm{diff}} - t\right)}{\int_{TP_1}^{TP_1+TP_{\mathrm{diff}}} f_{kernel}(t)\,\mathrm{d}t}, \tag{8.21}$$

where ${}_s\mathcal{P}_{kernel}$ is the sWeight of this event for this kernel class, and the integral in the denominator is of the unbiased decay time PDF over the accepted interval. The unbiased decay time PDF is not known, so an estimate of a single exponential convolved with a Gaussian is used initially, called $f_{0,kernel}(t)$. The lifetime parameter of the exponential is taken to be the mean of the decay-time data, sWeighted for this class, and the width of the Gaussian is taken to be the same as for prompt signal.

The biased PDF, constructed from the kernelised sPlots is divided by $f_{0,kernel}(A|t)$ to give the next estimate of the unbiased PDF. An iterative process is used to refine the unbiased PDF estimate such that,

$$f_{n+1,kernel}(t) = \frac{f_{kernel}(t, A)}{f_{n,kernel}(A|t)}. \tag{8.22}$$

For each step of the iteration the new unbiased PDF is compared to the previous one. When there is little change between successive iterations the process is deemed to have converged.

The kernels in decay time and $\ln(\chi^2_{\mathrm{IP}})$ are considered to be independent, with two bandwidths to describe the width of the kernels in each observable. The largest component for which this procedure is used is the combinatorial background. This approach relies on the assumption of a negligible correlation between the relevant observables. No significant correlation was observed for the combinatorial background or the misreconstructed backgrounds in simulated data.

8.9 Blinding

The real fitted lifetimes in the K^+K^- and $\pi^+\pi^-$ final states were hidden from the analysts ("blinded") during the course of setting up the fitting process and the systematic checks. This is done to ensure no analyst bias was introduced into the method. The parameters were multiplied by a "blinding" parameter at the interface between the numerical minimisation algorithm [71] and the fitter classes. The result was that unblinded parameters were fitted to the data but blinded parameters were returned to the user.

For all fits, the blinding parameter was a random scale, $b_{untagged}$, chosen according to a flat random distribution between 0.97 and 1.03. Flavour specific fit lifetimes were multiplied by a blinding factor $b_{D^0} = b_{untagged} + b_{tagged}$ for D^0 and $b_{\overline{D}^0} = b_{untagged} - b_{tagged}$ for $\overline{D}^0$, with b_{tagged} chosen according to a flat random distribution between -0.03 and 0.03. Such a method allows one to compare the results for systematics studies whilst keeping the true measured value unknown. The difference between measurements in the K^+K^- and $\pi^+\pi^-$ final states was unblinded during the course of the analysis by setting the final state specific blinding factors to be the same.

Table 8.1: Fit parameter results for the fit to D^0 decay-time and $\ln(\chi^2_{\mathrm{IP}})$ from one $D^0 \to K^+K^-$ sub-dataset with magnet polarity down and from the second run block. The symbols are explained in the text. The parameter units are shown where relevant.

Parameter	Fixed?	Value
$P(prom.\vert sig.)$	N	(0.91969 ± 0.00053)
τ	N	$(0.41325 \pm 0.00066)\,\mathrm{ps}$
$P(\chi 1\vert prom.)$	N	(0.9868 ± 0.0012)
$I_{\mu,1,prom.}$	N	(0.7926 ± 0.0048)
$S_{\mu,prom.}$	N	(0.1765 ± 0.0035)
$I_{\alpha_L,prom.}$	N	(1.0574 ± 0.0028)
$S_{\alpha_L,prom.}$	N	(-0.0624 ± 0.0022)
$\alpha_{R,1,prom.}$	N	(0.8898 ± 0.0050)
$I_{\mu,2,prom.}$	N	(2.976 ± 0.062)
$\alpha_{L,2,prom.}$	Y	0.9622
$\alpha_{R,2,prom.}$	N	(2.57 ± 0.28)
$P(sec.1\vert sec.)$	N	(0.9164 ± 0.0026)
$\tau_{1,sec.}$	Y	$0.33\,\mathrm{ps}$
$\tau_{2,sec.}$	N	$(0.959 \pm 0.014)\,\mathrm{ps}$
$\tau_{3,sec.}$	N	$(0.3261 \pm 0.0071)\,\mathrm{ps}$
g_0	N	(-3.997 ± 0.064)
g_1	N	$(0.997 \pm 0.020)[\mathrm{ps}]^{-1}$
g_2	N	(9.564 ± 0.085)
$p_{L,0}$	N	(1.49 ± 0.12)
$p_{L,1}$	N	$(0.047 \pm 0.023)[\mathrm{ps}]^{-1}$
$p_{L,2}$	N	$(-0.84 \pm 0.22)[\mathrm{ps}]^{-1}$
$p_{R,0}$	N	(-0.6486 ± 0.0075)
$p_{R,1}$	N	$(0.0138 \pm 0.0030)[\mathrm{ps}]^{-1}$
$p_{R,2}$	Y	$-8\ [\mathrm{ps}]^{-1}$
$P(prom.\vert \pi_s)$	N	(0.91610 ± 0.00087)
τ_{π_s}	N	$(0.4230 \pm 0.0012)\,\mathrm{ps}$

Chapter 9

Cross-checks and systematic uncertainties

Studies that validate the fit method are presented in this chapter. Data-driven techniques and simulated datasets are used to show that no significant bias is present in the measured value of A_Γ obtained from the method detailed in previous chapters, and using Eq. (2.48).

Firstly, the $K^-\pi^+$ control sample is studied to validate the method in Section 9.1, while a number of null tests on real and simulated data are presented in Section 9.2 and Section 9.3.

Tests for systematic biases inherent to the method are investigated in Section 9.4. Potential biases in the results due to the various backgrounds, from the acceptance determination and from other sources, are explored.

Finally, Section 9.5 shows a series of additional tests including splitting the data into disjoint samples by applying selection requirements to kinematic observables. The lack of significant variation in these sub-datasets shows that the value of A_Γ is not dependent on these selections.

9.1 Pseudo-A_Γ cross-check

Pseudo-A_Γ is defined as the value of A_Γ calculated from the D^0 and $\overline{D}^0$ lifetimes of the CF decay $D^0 \to K^-\pi^+$, which should be consistent with zero, as described in Section 2.5. Any statistically significant deviation from zero would indicate a bias in the fit method described in previous chapters.

Table 9.1: Measured values of $\tau(K^-\pi^+)$ for each of the 12 sub-datasets on which the individual fits are performed. The average lifetime is calculated as a weighted average with the sub-dataset size.

	MagDown		
$\tau(K^-\pi^+)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	414.36 ± 0.25	415.34 ± 0.21	415.52 ± 0.19
$\overline{D}^0$	414.02 ± 0.24	415.37 ± 0.21	415.23 ± 0.18
Average	414.18 ± 0.17	415.36 ± 0.15	415.37 ± 0.13
	MagUp		
$\tau(K^-\pi^+)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	414.00 ± 0.23	415.14 ± 0.19	415.39 ± 0.26
$\overline{D}^0$	413.82 ± 0.23	415.30 ± 0.18	415.61 ± 0.25
Average	413.91 ± 0.16	415.22 ± 0.13	415.50 ± 0.18

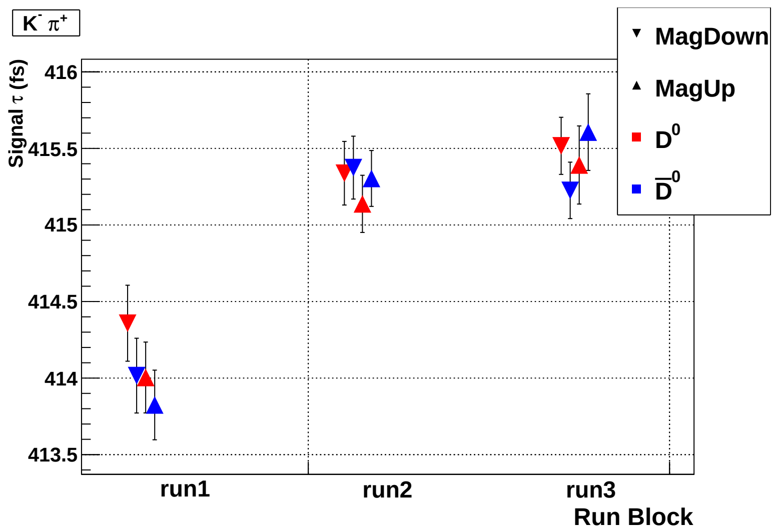


Figure 9.1: Measured values of $\tau(K^-\pi^+)$ for each of the 12 sub-datasets on which the individual fits are performed. The down arrows indicate a magnet polarity down sample and the up arrows indicate a magnet polarity up sample. A red marker indicates a measurement of the D^0 sample and a blue marker is a measurement from the $\overline{D}^0$ sample. The x-axis shows the run block, with an arbitrary offset applied to adjacent points so the errors can be seen.

Table 9.2: Measured values of $\tau(K^-\pi^+)$ when averaging over the two magnet polarities. The average lifetime is calculated as a weighted average with the sub-dataset size.

$\tau(K^-\pi^+)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	414.17 ± 0.17	415.23 ± 0.14	415.47 ± 0.15
$\overline{D}^0$	413.91 ± 0.17	415.33 ± 0.14	415.36 ± 0.15
Average	414.04 ± 0.12	415.28 ± 0.10	415.42 ± 0.11

The world average measurement of the D^0 lifetime in the decay $D^0 \to K^-\pi^+$ from [9] is, $\tau(K^-\pi^+) = (410.1 \pm 1.5)$ fs. Table 9.1 and Fig. 9.1 show the measured values of $\tau(K^-\pi^+)$ from the analysis in this thesis, for each of the 12 sub-datasets of the full $K^-\pi^+$ data. Table 9.2 shows the results when averaging over the magnet polarities. The averages are calculated by weighting with the total number of candidates in each sub-dataset, where only statistical uncertainties are shown. The total averaged result of (415.00 ± 0.06) fs is significantly higher than that of the world average, at the level of 3.3σ. There is also a discrepancy in the measured values between the first run block and the other two run blocks. Averaged values in the two magnet polarities and over both D^0 flavours are compatible. A χ^2 test of the compatibility of the 12 individual measured values yields a $< 1\%$ chance that these differences are caused by statistical fluctuations. The significant variations in the measured lifetimes are to be expected due to residual biases caused by selection and the challenging environment of hadron collisions. However, these biases are correlated for the D^0 and $\overline{D}^0$ sub-datasets, and hence while the absolute lifetimes are susceptible to systematic effects, the ratio of lifetimes is not significantly affected. Fig. 9.2 shows the calculated values of pseudo-A_Γ in the individual sub-datasets, where a good agreement between the measurements is evident. While the measured lifetime varies significantly between sub-datasets, the values of pseudo-A_Γ show no variation with run block or magnet polarity and have a χ^2 of compatibility of 70%.

The final value of pseudo-A_Γ is found to be $(-0.07 \pm 0.15) \times 10^{-3}$, which is consistent with the expected value of zero. These results show that the method does not account for all decay-time biasing effects present in the data, but the biases that affect the D^0 and $\overline{D}^0$ lifetime asymmetry are successfully removed. Thus such a method cannot be used to measure the D^0 lifetime accurately, but

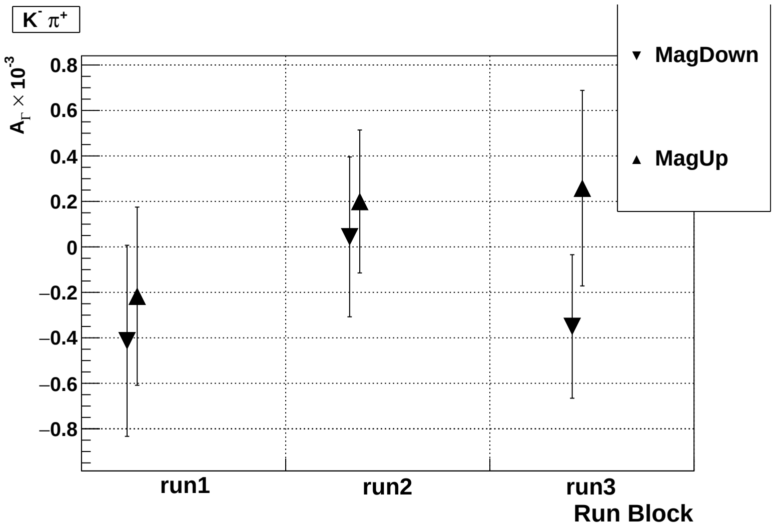


Figure 9.2: Calculated values of pseudo-A_Γ in the individual $K^-\pi^+$ sub-datasets. The down arrows indicate a magnet polarity down sample and the up arrows indicate magnet polarity up samples. The χ^2 probability of compatibility for these data is 70%.

provides unbiased measurements of A_Γ.

Fig. 9.3 shows the asymmetry in the number of D^0 and $\overline{D}^0$ candidates from the full $D^0 \to K^-\pi^+$ data, overlaid with the asymmetry of the fit model. Clearly there are significant decay-time dependent asymmetries in the data, but these are correctly modelled by the fit. These asymmetries are believed to originate from momentum dependent detection asymmetries for the π_s. The selection of soft pion candidates has a different efficiency depending on the pion charge and momentum. As the D^0 and π_s are related through the D^*, there is a relationship between the D^0 momentum and the π_s detection asymmetry. The selection also introduces a correlation between decay time and momentum, such that low momentum candidates need a larger decay time to be selected than that needed by high momentum candidates. Thus, there is a relationship between the D^0 lifetime and the π_s charge detection asymmetry. The swimming method parameterises D^0 momentum correlations with decay time and the method accounts for them in the fit, such that the link between the π_s charge detection asymmetry is removed and the correct value of A_Γ can be extracted.

The accuracy of the method is further demonstrated by comparing the asymmetry of the prompt signal component of the fit for each magnet polarity. The results from the two magnet polarities are compatible, as shown in Fig. 9.4 (left) and Table 9.3, which indicates consistent results. An additional demonstration of the robustness of the fit method is to calculate pseudo-A_Γ for the random π_s

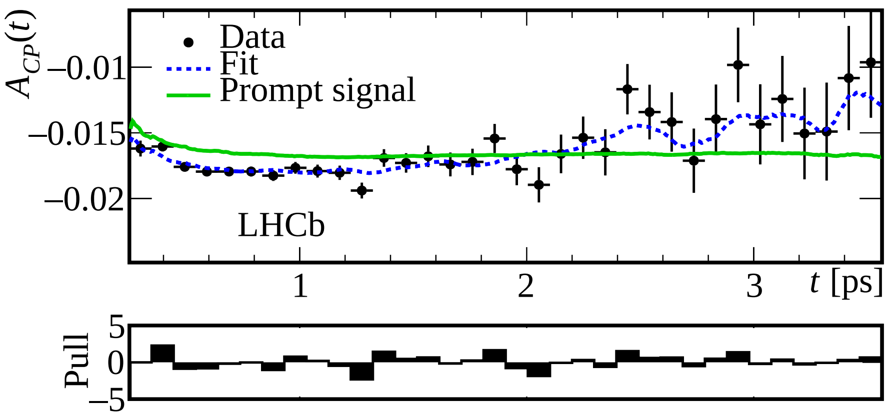


Figure 9.3: Asymmetry in the number of D^0 and $\overline{D}^0$ candidates, including backgrounds, as a function of decay time for the full $D^0 \to K^-\pi^+$ data overlaid with the asymmetry of the full fit model and the asymmetry of the prompt signal component from the fits. The flat pull distribution, which compares the data and the fit model, shows that the decay-time dependent asymmetries in the data are correctly modelled by the fit.

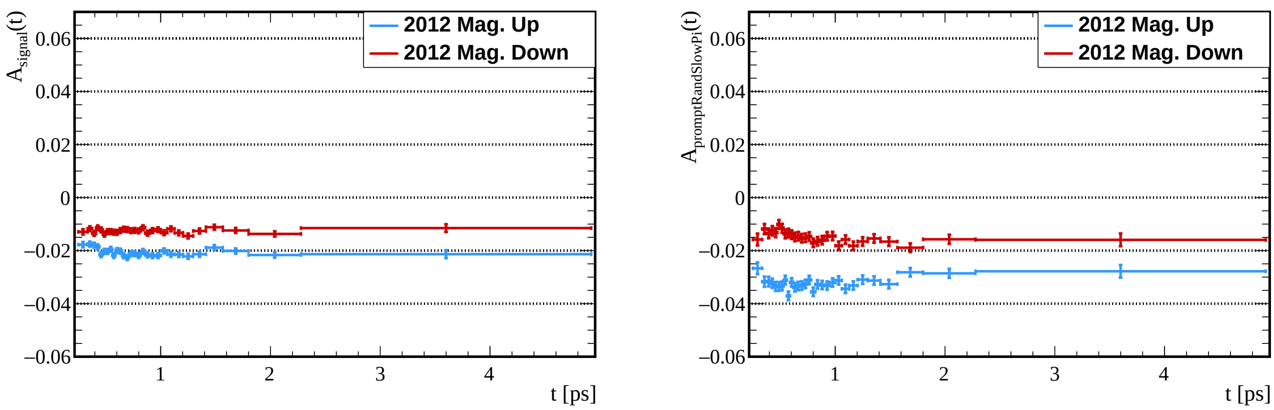


Figure 9.4: Asymmetry in the number of D^0 and $\overline{D}^0$ candidates as determined by the fit to $D^0 \to K^-\pi^+$ data as a function of decay time, comparing magnet polarities. The final bin extends up to 10 ps. The classes shown are (left) prompt signal and (right) prompt random soft pion background.

Table 9.3: The average values of pseudo-A_Γ for each magnet polarity, and their compatibility, for both prompt signal and random soft pion background.

Pseudo-A_Γ ($\times 10^{-3}$)	MagUp(fs)	MagDown(fs)	Difference(σ)
prompt signal	0.05 ± 0.21	-0.29 ± 0.20	1.15
prompt random π_s	-0.49 ± 0.53	0.87 ± 0.54	1.79

background, which should also be consistent with zero. As shown in Fig. 9.4 (right) the variation as a function of decay time of the asymmetry of this background is larger than that of signal. Nonetheless, Table 9.3 shows that the expected values of pseudo-A_Γ are obtained, and that they are compatible between magnet polarities. This demonstrates that the fit method is robust against momentum dependent detection asymmetries.

9.2 Randomised flavour tag null tests

The fits to K^+K^- and $\pi^+\pi^-$ data are verified by randomising the D^0 and $\overline{D}^0$ flavour tags. A uniform random number between 0 and 1 is chosen. If the number is less than or equal to the fraction of D^0 events in the dataset then it is tagged as a D^0 event. This fraction is found to be 0.495. If the chosen random number is above this value then the candidate is given a $\overline{D}^0$ tag. In this way, any possible asymmetry between the D^0 and $\overline{D}^0$ lifetimes is removed from the dataset and the fit should return an A_Γ value of zero on average. For each final state 100 such tests are run with a different seed used to generate the random numbers in each case. Table 9.4 shows the results for both final states and Fig. 9.5 shows the pull distributions of the measured values for the K^+K^- and $\pi^+\pi^-$ samples. The pull distribution is the difference between the measured values and the expected values divided by the measurement uncertainties. They are consistent with a normal distribution centered at zero, which indicates that the results are unbiased. Thus the method as outlined in previous chapters produces unbiased measurements of $A_\Gamma(K^+K^-)$ and $A_\Gamma(\pi^+\pi^-)$.

Table 9.4: Results of null tests on K^+K^- and $\pi^+\pi^-$ data with randomised flavour tags. The average A_Γ value from this test is zero showing that the method works as expected. The A_Γ mean bias is the difference between zero and the measured values of A_Γ for each test, while the A_Γ pull mean (relative bias) is the mean value of the difference divided by the error for each test. The χ^2 probability is from the fit of a normal distribution to the pull distribution.

Final State	A_Γ mean bias($\times 10^{-3}$)	A_Γ pull mean	$P(\chi^2\vert\mathrm{ndf})$
$\pi^+\pi^-$	-0.012 ± 0.075	-0.015 ± 0.095	52.3%
K^+K^-	0.038 ± 0.047	0.083 ± 0.103	10.4%

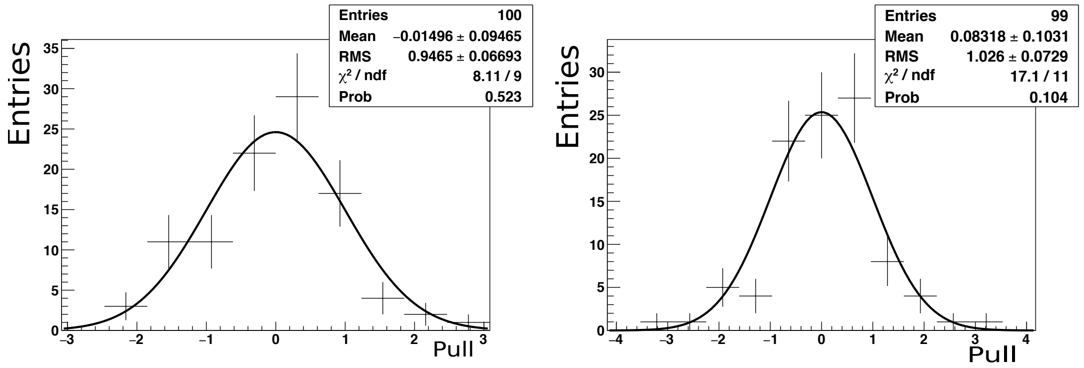


Figure 9.5: Pull distribution of the calculated values of (left) $A_\Gamma(\pi^+\pi^-)$ and (right) $A_\Gamma(K^+K^-)$ relative to the expected value of zero, using a dataset with approximately 100 different seeds used to randomise the D^0 flavour tag. A χ^2 test for consistency with a Gaussian with $\mu = 0$ and $\sigma = 1$ is shown.

9.3 Control toy studies

A further verification of the accuracy of the method is obtained via fits to simulated pseudo-data, known as "toy" data. These data are generated using the same PDFs as are used to fit the data for each of the final states. The PDF parameter values are taken from the fits to data. One exception is the generated lifetime value that is set to 406.6 fs, which is approximately the D^0 lifetime when decaying to a *CP* eigenstate. Running the fits on toy data can demonstrate the internal consistency of the method. Additionally, potential biases and correlations can be added to the toy data to check their effects on the measured values; this is further discussed in the following sections. The toy data are randomly divided up between D^0 and $\overline{D}^0$, magnet up and down, and three run blocks with fractions roughly equal to those of the data in each case. In particular, there is a 0.5% difference in statistics for D^0 and $\overline{D}^0$, as is seen in data.

Approximately 200 such toy datasets were generated for each final state, each with a total of 4.8 million candidates and the full fit run on each. The distribution of the fitted values and their pulls are examined for bias. If the fitted values are unbiased and the statistical uncertainties accurate, then the distribution of pulls should be consistent with a Gaussian with $\mu = 0$ and $\sigma = 1$. Fig. 9.6 shows the pull distributions for the K^+K^- and $\pi^+\pi^-$ final states, averaged over the six sub-datasets. The corresponding mean biases are quoted in Table 9.5. The χ^2 tests of compatibility with normal distributions indicate that the method is unbiased for both final states.

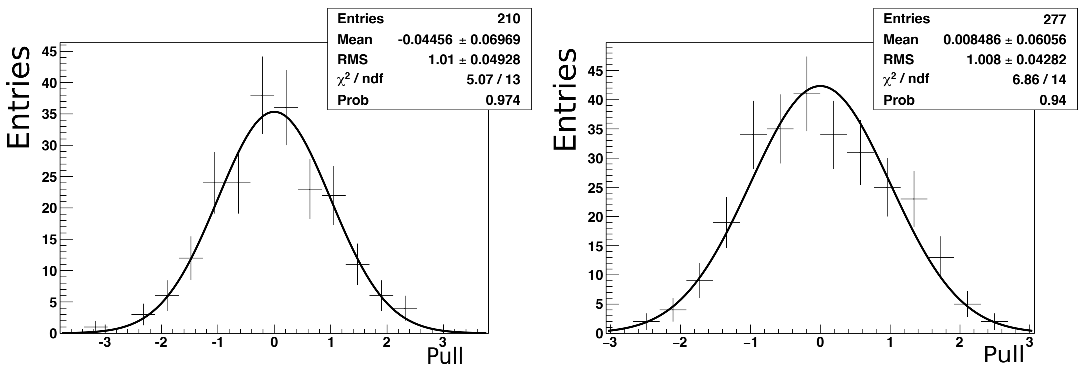


Figure 9.6: Pull distribution of the calculated values of (left) $A_\Gamma(\pi^+\pi^-)$ and (right) $A_\Gamma(K^+K^-)$ relative to the expected value of zero, using control toy datasets. A χ^2 test for consistency with a Gaussian with $\mu = 0$ and $\sigma = 1$ is shown.

Table 9.5: Results from the control toy tests for $A_\Gamma(K^+K^-)$ and $A_\Gamma(\pi^+\pi^-)$. The mean bias is the measured value of A_Γ for all datasets relative to the expected value of zero, while the A_Γ pull mean (relative bias) is the mean value of the pull distribution.

Final State	A_Γ mean bias($\times 10^{-3}$)	A_Γ pull mean
K^+K^-	0.055 ± 0.041	0.008 ± 0.061
$\pi^+\pi^-$	-0.056 ± 0.088	-0.045 ± 0.070

9.4 Systematic uncertainties

A number of tests are carried out to evaluate the size of the systematic uncertainties and to check the stability of the fit results. This section describes the procedure for determining the systematic uncertainties and provides a summary of the results obtained, which are given in Table 9.6.

The systematic uncertainties are evaluated by a combination of varying the fit configuration and by adding various effects to toy datasets. In the former case the alternative configurations are applied in turn to the method and values of A_Γ determined. The corresponding systematic uncertainty is evaluated by taking half the amplitude of the maximum variation of the result between the alternate configurations. If the alternate configurations are systematically higher or lower than the default fit result, then the full amplitude of the maximum variation is taken. If there is only one alternate value with which to compare, the difference between this value and the default value is taken as the systematic uncertainty.

In the latter case, toy datasets are produced with various added effects and

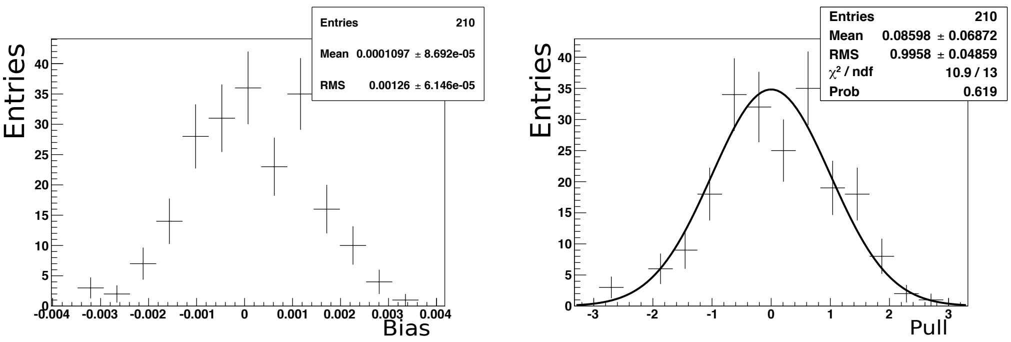
Table 9.6: Summary of systematic uncertainties for the K^+K^- and $\pi^+\pi^-$ final states. All uncertainties are given as multiples of 10^{-3}.

Effect ($\times 10^{-3}$)	$A_\Gamma(K^+K^-)$	$A_\Gamma(\pi^+\pi^-)$
combinatorial background	± 0.02	± 0.10
secondary contamination	± 0.08	± 0.08
correlations	± 0.02	± 0.07
low decay-time mismodelling	± 0.03	± 0.05
random soft pion background	± 0.02	± 0.03
turning point accuracy	± 0.03	± 0.02
decay-time resolution	± 0.01	± 0.01
specific backgrounds	± 0.01	—
total systematic uncertainty	± 0.10	± 0.16
total statistical uncertainty	± 0.46	± 0.79

fitted with the default fit configuration. As with previous toy tests, 200 sets of data are generated with 4.8 million events in each. In some cases 300 sets were generated. The systematic uncertainty is assigned by first considering the larger of either the relative bias of the test or the difference between the relative bias and that of the control toy test. Here the relative bias refers to the mean value of the pull distribution. The standard test has a relative bias of 0.008 for K^+K^- and -0.045 for $\pi^+\pi^-$ as in Table 9.5. The assigned systematic is the largest of the two possible biases described above, scaled by the statistical uncertainty on A_Γ, which accounts for the different sample sizes between the real and the toy datasets. The CPU requirements for generating the same number of toy events as data events is prohibitive. The amount of statistics generated is considered to be sufficient.

9.4.1 Combinatorial background

Contamination from combinatorial background is the largest systematic effect in the $\pi^+\pi^-$ final state as it is a relatively large background. The impact of combinatorial background is studied with several toy tests. The first test varies the $m(D^0)$ and Δm description by generating data with alternative PDFs. A linear distribution is used to generate the $m(D^0)$ distribution and a modified threshold function, given in Eq. (7.5), is used to generate the Δm background.



(a) (b)

Figure 9.7: Results of $A_\Gamma(\pi^+\pi^-)$ from toy fits with an alternate $m(D^0)$ and Δm fit configuration as described in the text. Shown are, (a) the calculated values of $A_\Gamma(\pi^+\pi^-)$ and (b) their pull, relative to the generated value of zero, on each individual sub-dataset. A χ^2 test for consistency with a Gaussian with $\mu = 0$ and $\sigma = 1$ is shown for the distribution of pulls.

The second test uses a varied time distribution for the combinatorial background; the default toy experiments use a single exponential with a lifetime of 300 fs, while this variation uses a combination of a short-lived, a medium-lifetime, and a long-lived component.

Fig. 9.7 shows the pull plots for the first toy test on the $\pi^+\pi^-$ final state, which gives the largest systematic uncertainty of 0.10×10^{-3} from a relative bias of 0.086, when compared to the control toy test relative bias. The second toy test has a relative bias of 0.075. A third toy test is carried out for this final state due to the large fraction of combinatorial events in the dataset. This test doubles the fraction of combinatorial events in the toys. The resulting relative bias is -0.055 giving a systematic uncertainty of 0.04×10^{-3}. The largest of these is assigned as the systematic uncertainty on the final result.

For K^+K^- the second test yields the largest systematic uncertainty of 0.02×10^{-3} from a relative bias of 0.042. The first test has a relative bias of 0.015 giving a smaller systematic uncertainty.

9.4.2 Secondary contamination

Residual contamination from secondary decays is a large systematic in both final states, as it is difficult to separate from the prompt signal events, especially at low decay-times where the distributions overlap.

The impact of secondary contamination is assessed in several ways. The first

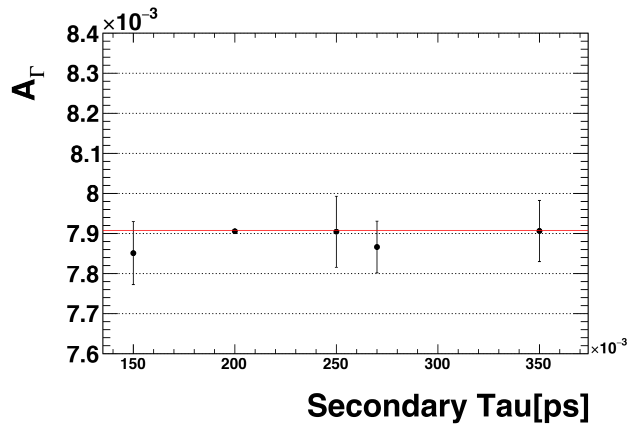
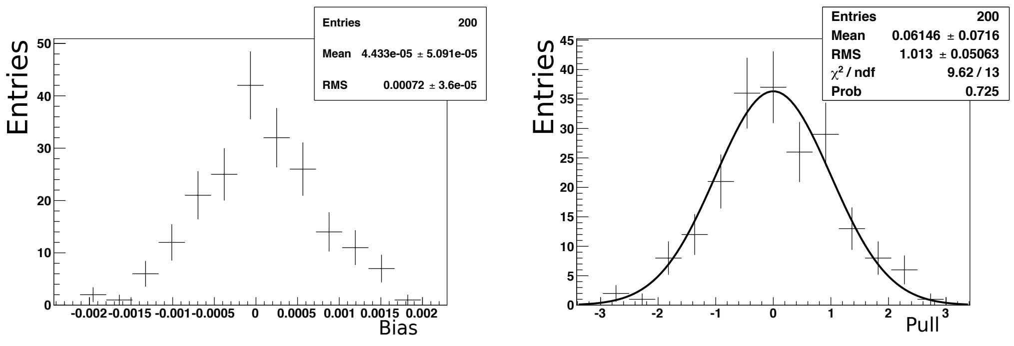


Figure 9.8: Variation of blinded $A_\Gamma(K^+K^-)$ as the first lifetime parameter of the secondary decay-time component is fixed to different values in the fit. The default blinded A_Γ value, with the parameter fixed to 0.33 ps is shown with a red line. The errors shown are the quadratic subtraction of the default error from the errors returned by the alternate fits.

test is a set of toy studies with a varying fraction of secondary decays, starting from a nominal value of about 5% of the total dataset, up by a factor of two and down by a factor of 0.5. The secondaries make up about 7% of the signal class, with prompt signal making up the rest. These ranges are large enough to cover a difference seen between the evolution of the secondary fraction with decay-time in the simulation and the fit results. A second test measures the sensitivity to the first lifetime parameter of the secondary decay-time distribution, presented in Eq. (8.14), that is fixed by default.

No significant deviations or biases are seen in any of the tests, nonetheless a systematic is assigned for each final state. For the K^+K^- final state the varying of the first lifetime parameter gives the largest systematic error of 0.08×10^{-3}, shown in Fig. 9.8. The toy studies that double and halve the fraction of secondary events give systematic uncertainties of $0.04{\times}10^{-3}$ and $0.02{\times}10^{-3}$, respectively.

For the $\pi^+\pi^-$ final state the toy test that halves the secondary fraction gives the largest systematic uncertainty. Fig. 9.9 shows that the relative bias of the test is 0.061, which gives a systematic uncertainty of $0.08{\times}10^{-3}$ when compared to the control test. The test that varies the first lifetime parameter for $\pi^+\pi^-$ gives a systematic error of 0.06×10^{-3}. Finally, the other toy test that doubles the fraction of secondary events gives a systematic uncertainty of $0.05{\times}10^{-3}$.



(a) (b)

Figure 9.9: Results of an $A_\Gamma(\pi^+\pi^-)$ toy study with the default fraction of secondary events halved. Shown are, (a) the calculated values of $A_\Gamma(\pi^+\pi^-)$ and (b) their pull, relative to the generated value of zero, on each individual sub-dataset. A χ^2 test for consistency with a Gaussian with $\mu = 0$ and $\sigma = 1$ is shown for the distribution of pulls.

9.4.3 Correlations between variables

This section discusses the correlations between variables in the data that are not accounted for in the fit method. Tests of the systematic effects caused by omitting them are carried out with generated toy data. Samples of simulated events were used to parameterise the correlations such that they could be added to toy data.

The fit method assumes that the variables of the first stage fit are independent of those in the second stage, however this is not the case for prompt $\ln(\chi^2_{\rm IP})$ and $m(D^0)$. The dependence can be seen in Fig. 9.10(left), which shows the mean value of $\ln(\chi^2_{\rm IP})$ in bins of $m(D^0)$ for simulated prompt data. In the fit method these dependencies are not accounted for. Similarly, $\ln(\chi^2_{\rm IP})$ for secondary events in data depends on the decay time and TP_1. In particular, the sigmoid function that describes the decay-time dependence of the $\ln(\chi^2_{\rm IP})$ mean has a TP_1 dependence. The dependence of the zeroth sigmoid function parameter is shown against TP_1 in Fig. 9.10(right). A significant correlation is seen between the $\ln(\chi^2_{\rm IP})$ mean and TP_1. The fit method parameterises the mean and tail parameters of the secondary $\ln(\chi^2_{\rm IP})$ PDF as a function of decay-time only. To test these effects toy data are generated with the two correlations of Fig. 9.10 added.

The correlations alone will not affect the value of A_Γ unless the magnitude of the effect is different for D^0 and $\overline{D}^0$. There are two scenarios in which this is possible. The first is if the value of A_Γ is non-zero. In this case, if the bias created by ignoring the correlations in the fit model is lifetime dependent, then

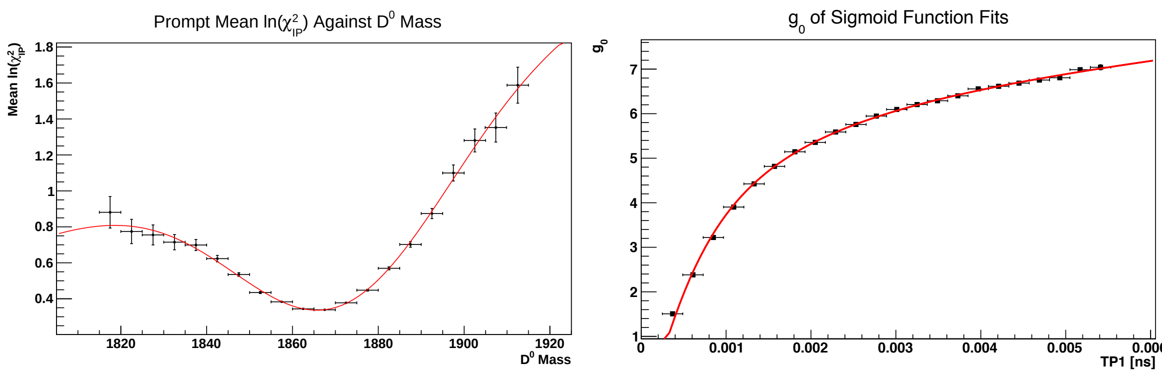


Figure 9.10: Profile plots from fits to simulated $K^-\pi^+$ data for (left) the prompt $\ln(\chi^2_{\mathrm{IP}})$ mean parameter in bins of $m(D^0)$ and (right) the zeroth parameter of the sigmoid function of Eq. (8.17), which describes the secondary $\ln(\chi^2_{\mathrm{IP}})$ mean, in bins of TP_1.

a difference in the D^0 and $\overline{D}^0$ lifetimes will bias the value of A_Γ. The second scenario is if the bias created by the correlations is dependent on the sample size. In this case the effect may not cancel in the lifetime ratio and the value of A_Γ may be biased.

Two sets of toy data are generated and fitted to test these effects. The first set has a non-zero value of A_Γ. One sample has a generated signal lifetime of 406.1934 fs, while the other has a signal lifetime of 407.0066 fs giving an A_Γ value of $1{\times}10^{-3}$. This is more than 2σ greater than the current world average. The second set is generated with a large difference in statistics between D^0 and $\overline{D}^0$. In this test the percentage of D^0 events is reduced to 47.5%.

For the first test, a significant bias is observed between the generated values of the signal lifetimes and the results from the fits. This is due to the difference between the nominal fit configuration and the generated toy data. In the toy fits the extra secondary correlations cause the secondary $\ln(\chi^2_{\mathrm{IP}})$ and decay-time components to have different shapes compared to the nominal PDFs. This changes the fitted prompt signal lifetime. However, the bias in the samples with different generated lifetimes is the same. No bias is observed between the calculated value of A_Γ and the generated value for K^+K^- or $\pi^+\pi^-$. The lifetime biases cancel when A_Γ is calculated as the biases in the D^0 and $\overline{D}^0$ samples are the same. None of the tests showed significant deviations in the relative biases, however systematic uncertainties are assigned to account for possible effects that may have been evident with larger toy dataset statistics. The relative bias on A_Γ due to this test is 0.052, scaling by the K^+K^- statistical uncertainty gives a systematic uncertainty of 0.02×10^{-3}.

For the second test with the K^+K^- final state, the relative bias is -0.046, which also gives the systematic uncertainty as 0.02×10^{-3}. For the $\pi^+\pi^-$ final state, the second test gives the largest relative bias of 0.042, resulting in a systematic uncertainty of 0.07×10^{-3}.

It was also seen that the secondary $\ln(\chi^2_{\rm IP})$ has a dependence on $m(D^0)$, however this is a small effect compared to the TP_1 and decay-time dependence. Due to a lack of statistics in the simulation sample it is not possible to include more variables in the secondary mean $\ln(\chi^2_{\rm IP})$ parameterisation. A small correlation is also seen between the secondary $\ln(\chi^2_{\rm IP})$ and the turning point difference $\mathrm{TP}_{\rm diff}$. Again, due to lack of MC statistics it is not clear if this is a residual effect of the correlation between $\ln(\chi^2_{\rm IP})$ and TP_1 or a separate effect. For the prompt simulation sample, fits to the decay-time dependence of the $\ln(\chi^2_{\rm IP})$ mean in bins of TP_1 suggest a small correlation, but it is a small effect compared to the $\ln(\chi^2_{\rm IP})$ and $m(D^0)$ relationship, shown in Fig. 9.10(left). Toy tests including the above small relationships were done and no significant relative biases were observed. No further systematic uncertainties are assigned for these effects.

9.4.4 Low decay-time mismodelling

There is an evident difference between the fit model and the data in the low decay-time region, as in Fig. 8.9(left). By using toy studies and carefully examining the pull distributions between the generated data and the total fit results, it is found that the neglected correlations described in Section 9.4.3 are the largest contributors to the mismodelling. An over-estimation of the $K^-\pi^+\pi^0$ kernel background at low decay times also contributes for the K^+K^- final state, a systematic test for which is discussed in Section 9.4.8. However, the effect below about 0.75 ps is still not well understood.

The asymmetry plots, as shown in Fig. 9.3 for $K^-\pi^+$, show that the fit model describes the data well. This is only possible if the decay-time mismodelling of the fit is the same for the D^0 and $\overline{D}^0$ sub-datasets, and thus cancels in the final calculation of A_Γ. A systematic uncertainty is assigned to quantify the degree to which the effect cancels. An arbitrary decay-time offset that reproduces the mismodelling in the K^+K^- final state is used to assess the systematic. The offset is parameterised in terms of decay time and 330 toy data sets are generated and compared with an equal number without the bias. The results for the measured A_Γ and lifetime are shown in Table 9.7.

Table 9.7: Results of toy studies with and without the low decay-time offset for the K^+K^- final state. Shown are the relative biases, or the means of the pull distributions.

	relative bias	
	no offset	with offset
A_Γ	0.070 ± 0.059	0.007 ± 0.059
τ	0.301 ± 0.059	0.121 ± 0.059

The relative bias is taken as the difference between the toy tests with and without the bias. Scaling by the relevant statistical uncertainty, the systematic associated to the K^+K^- final state for this effect is 0.03×10^{-3} and for the $\pi^+\pi^-$ final state is 0.05×10^{-3}. Further toy studies of this effect, which change the shape of the effect by introducing a non-zero value of A_Γ were carried out, but no bias was observed.

9.4.5 Random soft pion background

The background from D^0 candidates that are associated with a random soft pion is assumed to follow a single exponential decay-time distribution. Naïvely, one would expect this lifetime to reflect the lifetime of a *CP* eigenstate, *i.e.* about 407 fs. This however ignores potential correlations with other backgrounds that have not been the focus of these studies. The lifetime parameter associated with the random soft pion background is the parameter with the largest correlation with the signal lifetime and therefore deserves closer attention.

Two tests are carried out to assess the size of potential systematic effects. Since this background contains true D^0 candidates, its shape in $m(D^0)$ is the same as the signal peak in the default fit. In the first test the random soft pion shape in $m(D^0)$ is made fully independent of the signal peak. The second test parameterises the random soft pion lifetime as a multiple of the signal lifetime in the fit. This multiplicative factor is allowed to float in the fit.

No statistically significant deviation is seen when comparing the test results to the default values of A_Γ; nonetheless, a systematic is still assigned. The first test yields a variation of 0.02×10^{-3} for K^+K^- and 0.01×10^{-3} for $\pi^+\pi^-$. The variation in the second test is negligible for K^+K^- and has a value of 0.03×10^{-3} for $\pi^+\pi^-$. Thus, the systematics assigned for K^+K^- and $\pi^+\pi^-$ are 0.02×10^{-3} and 0.03×10^{-3}, respectively.

9.4.6 Turning point accuracy: shift and scaling of TP positions

Two tests assess potential inaccuracies in the determination of the position of the acceptance TPs. One test is used to judge the impact of potential scale differences between the measured decay time and that used to determine the TP position. The second test assesses the impact of a constant bias to the TPs. As both tests affect only the TPs and not the measured decay time of the event, they test a worst case scenario in which a systematic effect is present only in the method of determining the TPs.

In principle, no scale inaccuracy is expected for the TP positions as their associated decay time is evaluated in the same way as the real measurement of the decay time. Nevertheless, a scale factor of up to 1.0001 is applied to all turning points to mimic unexpected effects. Similarly, constant offsets to the turning points are not expected to occur through a particular mechanism. Again, their impact is tested to assess the potential of unaccounted-for effects. The size of the largest bias in the test is 0.005 mm on the position of each turning point. The values used for these tests are motivated by studies of the TP resolutions.

For $\pi^+\pi^-$ the turning point scale test gives a systematic of 0.02×10^{-3}, while the other test gives a value of 0.002×10^{-3}. In K^+K^- the turning point bias causes the largest variation leading to an uncertainty of 0.03×10^{-3}, while the TP scale test gives 0.01×10^{-3}.

9.4.7 Variation of decay-time resolution

The decay-time resolution is expected to have a negligible effect on the result given the overall excellent resolution of the LHCb detector. The fit assumes a fixed single Gaussian resolution with a width of 50 fs. This resolution is varied from 30 fs to 100 fs. In addition, a triple Gaussian resolution function, taken from a study of $B^+ \to J/\psi K^+$ decays, is used to test the impact of a small number of events that have much larger resolutions [82]. The values used are; $\sigma_1 = 32$ fs, $\sigma_2 = 65$ fs, $\sigma_3 = 380$ fs, with fractions $f_1 = 0.714$, $f_2 = 0.282$, and $f_3 = 0.004$, respectively [82].

For both final states the variation of the single Gaussian resolution gives a systematic uncertainty of 0.01×10^{-3}. Using the triple Gaussian resolution function has no significant systematic effect on the measured values of A_Γ.

9.4.8 Specific backgrounds

The effect of wrongly accounting for specific backgrounds in the K^+K^- final state is tested by changing the unbiased decay-time distribution for the specific background components in the fit. This is usually calculated in an iterative manner using a single exponential as the initial estimate, as presented in Section 8.8.

The test involves trialling different initial mean lifetimes in the exponential PDF. Three mean lifetimes are trialed for each background. The first two values are taken from fits to simulation under various conditions, and the third is chosen to be comparable to the mean value of the likelihood weighted decay-time data. The chosen values are 550 fs, 740 fs and 1200 fs for the $K^-\pi^+\pi^0$ background and 520 fs, 860 fs and 1250 fs for the D_s^+.

The variations are applied together in pairs for the two channels (i.e. 550 fs and 520 fs are applied together, 740 fs and 860 fs and so on). The alternate fit configurations do not significantly change the value of $A_\Gamma(K^+K^-)$. A systematic uncertainty of $0.01{\times}10^{-3}$ is assigned from these tests.

A third test fixes the specific background fractions, $P(K^-\pi^+\pi^0)$ and $P(D_s^+)$, to plus and minus one of the standard deviation returned by the fit. The fractions are reduced to values of 0.0253 and 0.004 for $K^-\pi^+\pi^0$ and D_s^+, respectively, in the first test and increased to 0.0287 and 0.0054 in the second test. The deviation from the nominal result due to this test is entirely negligible.

The final test changes the initial kernel bandwidth used by scaling Eq. (6.12) by 0.75, 1.25 and 1.5. A systematic variation of 0.01×10^{-3} is observed but is not significant.

9.4.9 Track reconstruction efficiency

The inefficiency of detecting tracks that originate far from the beamline is corrected with a per-track correction that was discussed in Section 8.3.3 and included in the fit method in Section 8.6. This correction is not used in the default fit, but is included to test for any systematic bias from ignoring the effect. The change in the value of A_Γ due to including this correction is of the order of 10^{-7} for the $\pi^+\pi^-$ final state. Thus, no systematic is assigned for the track reconstruction efficiency.

9.5 Additional cross-checks

Number of turning points

As mentioned in Section 8.5 the maximum number of TPs for a candidate is restricted to two in order to keep the calculation of the acceptance PDFs manageable. Relatively few candidates are removed by this cut compared to the sample sizes. To check that no bias is introduced, the fits were run with the maximum number of turning points set to four. In this case two top hats were potentially used in the normalisation of the decay time PDF, but the TP PDFs $f(A)$, were still restricted to a maximum of two TPs. The fit results were consistent with the default value of A_Γ for K^+K^- and $\pi^+\pi^-$.

Multiple candidates

Events with multiple true signal decays can occur but do so rarely. Most cases of multiple candidates in a single event are due to several soft pion candidates being associated with a single D^0 candidate. These are well separated using the Δm distribution. As mentioned previously, one candidate is chosen at random in multiple candidate events. The impact of this selection is measured by using a different random number generator seed. The correlation between these datasets is the square root of the ratio of events shared between them. This is estimated from the number of multiple candidate events in the dataset. For the $\pi^+\pi^-$ final state, the difference from the final result of this test and the nominal value, with an uncorrelated error is $(0.44 \pm 0.29) \times 10^{-3}$. The probability of this difference being a fluctuation is 13.4% and is deemed to be acceptable. No variation is seen in the K^+K^- tests.

A second test removes all events with multiple candidates and reruns the analysis. Many of the events in this test are shared with the default dataset so the correlations are considered as in the previous test. No variation is seen in this test for either final state.

Dataset splits

The internal consistency of the results is tested by splitting the datasets in various ways. The data were split by: D^0 momentum and transverse momentum, the pseudorapidity η and the angle ϕ, the number of PVs in the event and the L0 trigger decision (TIS/TOS and TOS/not TOS). One such split for D^0 transverse

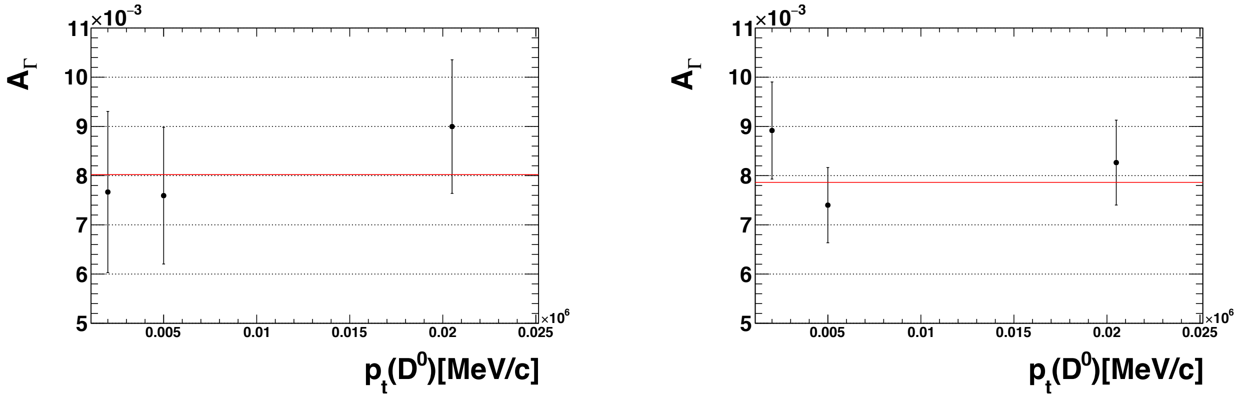


Figure 9.11: Variation of the blinded A_Γ results when splitting the $\pi^+\pi^-$ (left) and the K^+K^- (right) data by D^0 transverse momentum. The data points mark the centre of each bin and the default value is shown with a red line.

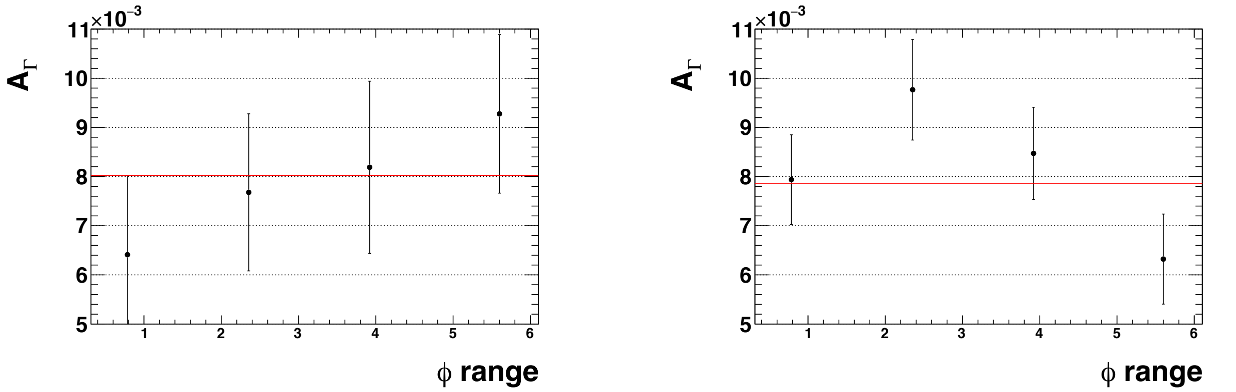


Figure 9.12: Variation of the blinded A_Γ results when splitting the $\pi^+\pi^-$ (left) and the K^+K^- (right) data by D^0 ϕ. The data points mark the centre of each bin and the default value is shown with a red line.

momentum is shown in Fig. 9.11. All data splits show no variation in the value of A_Γ, except the K^+K^- sample split by ϕ in which there is some deviation between datasets. However, with a χ^2 probability of 8% it is considered to be acceptable, as shown in Fig. 9.12(right).

The data were also split by soft pion χ^2_{IP}, PIDμ and PIDe. Again, no inconsistencies were seen in these data splits. To check for possible effects of correlation between random soft pion kinematics and decay time, the data are split by the fiducial cut decision, which vetos soft pions that are too close to the beamline or close to the outer edge of the detector fiducial volume. In theory this should be accounted for by the swimming algorithm when determining the VELO geometry acceptance, and included in the fit as part of the acceptance function.

No variation is seen between the two samples.

A final split was done with the pions' χ^2_{IP} for the $\pi^+\pi^-$ final state. No inconsistency was seen in the results from these data samples.

Chapter 10

Results and conclusions

10.1 Fit results

The fit projections for the $\pi^+\pi^-$ sub-dataset with magnet polarity down, from run block two and with D^0 flavour, are shown for the first fit stage to $m(D^0)$ and Δm in Fig. 10.1 and for the second fit stage to decay-time and $\ln(\chi^2_{\text{IP}})$ in Fig. 10.3. The $m(D^0)$, Δm and $\ln(\chi^2_{\text{IP}}(D^0))$ fits show an excellent fit quality, which is an impressive achievement given the large number of events. The decay-time projection has a discrepancy at low decay times for which systematics were assigned in Section 9.4.4. These fits are sufficient for the measurements of ratios of effective lifetimes as shown by the cross-checks and systematic studies reported in Chapter 9. Deviations between the data and the fits for D^0 and $\overline{D}^0$ are very similar leading to their cancellation in the final asymmetry calculations.

The fits for the K^+K^- final state with magnet polarity down, run block two and D^0 flavour were shown in Figs. 7.2 and 8.9. The fit quality is good, but again there is a difference between the fit and the data at low decay times.

The asymmetry projections of the fit results, shown in Fig. 10.2 for the $\pi^+\pi^-$ and K^+K^- final states, show that the asymmetries of the data are well modelled in the fit. To measure the agreement of the fit and the data, the χ^2 probability is calculated. It is not clear what the number of degrees of freedom of this projection of the final fit should be, so the number of bins in the asymmetry plot is used as an estimate. This gives a probability of 94.7% for $\pi^+\pi^-$ and 61.4% for K^+K^-. For the $K^-\pi^+$ final state the asymmetry projection was shown in Fig. 9.3, and has a χ^2 probability of 99.6%

The $K^-\pi^+$ fits for the same sub-dataset are presented in Figs. 10.4 and 10.5. The fit quality is poor in all variables, which reflects the large size of the dataset.

Despite these discrepancies the method yields the correct result for pseudo-A_Γ as shown in Section 9.1.

The resulting fit parameters for the K^+K^- final state fits can be found in Tables 7.2 and 8.1. The corresponding tables for the $\pi^+\pi^-$ and $K^-\pi^+$ final states can be found in Appendix B.

10.2 A_Γ results

Shown in Tables 10.1 and 10.2 are the values of $\tau(D^0)$ and $\tau(\overline{D}^0)$ split by magnet polarity and run block for both final states. Significant variation is seen in the individual lifetimes, however, this does not affect the measured values of A_Γ in the sub-datasets, as shown in Figs. 10.6 and 10.7.

The final unblinded values of A_Γ are,

$$\begin{aligned} A_\Gamma(K^+K^-) &= (-0.03 \pm 0.46 \pm 0.10) \times 10^{-3}, \\ A_\Gamma(\pi^+\pi^-) &= (\ \ 0.03 \pm 0.79 \pm 0.16) \times 10^{-3}, \end{aligned}$$

where the first uncertainty is statistical and the second is systematic. Thus the measured values for the two final states are found to be compatible with each other and with zero, showing no evidence for CP violation.

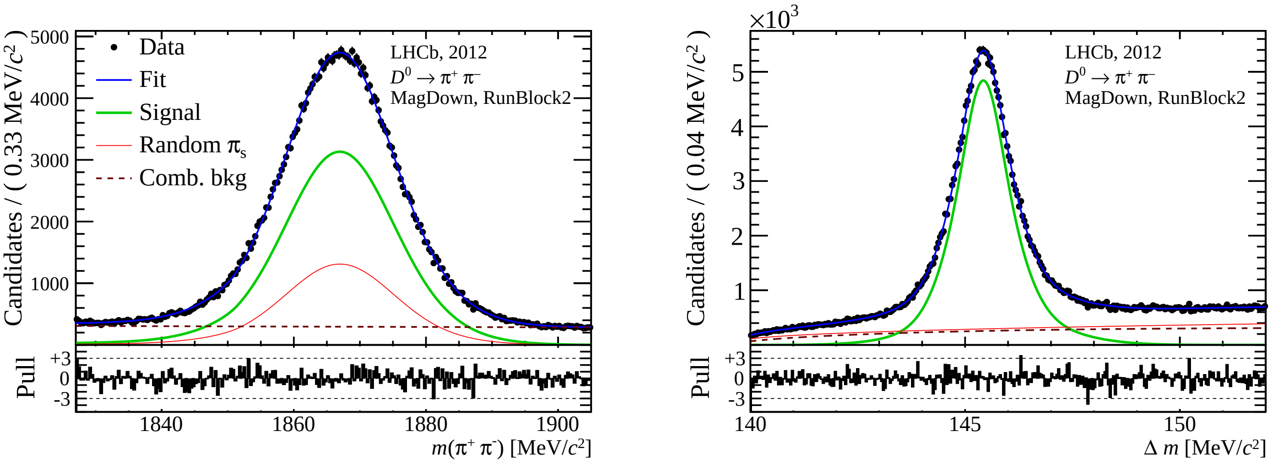


Figure 10.1: Fit projections of (left) $m(D^0)$ and (right) Δm for the $\pi^+\pi^-$ final state in the magnet polarity down, run block two and D^0 flavour sub-dataset.

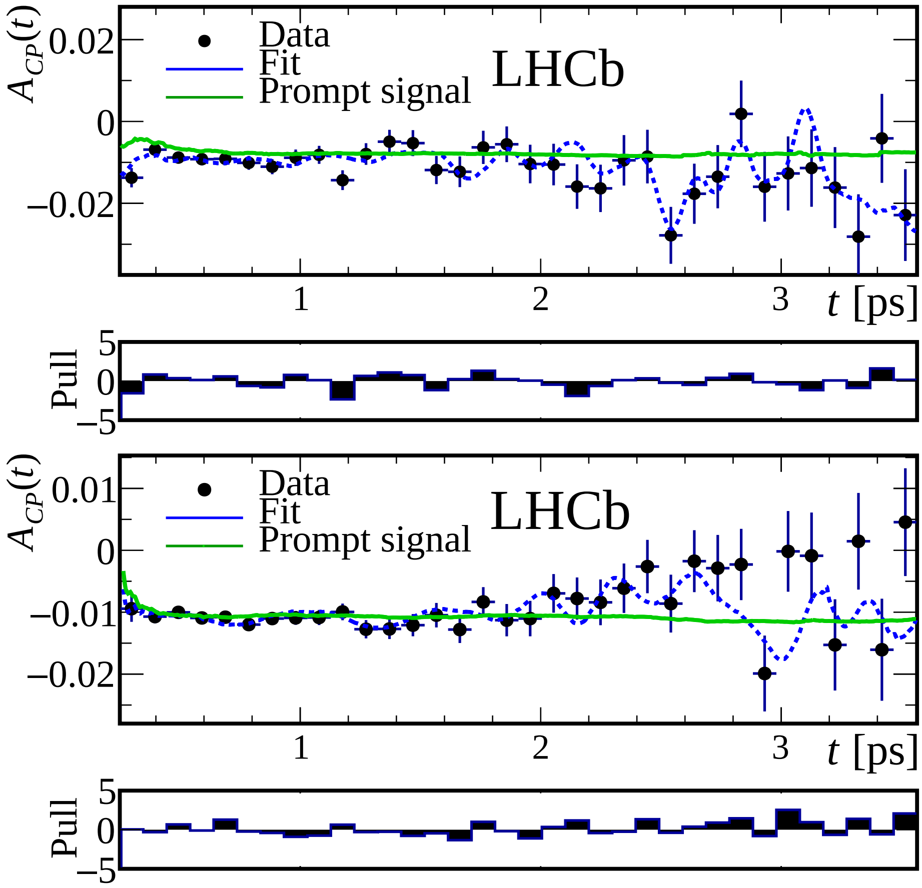


Figure 10.2: Asymmetry of the $\overline{D}^0$ to D^0 candidates, including backgrounds, overlaid with the asymmetry projection of the fit and of the prompt signal component for (top) $\pi^+\pi^-$ and (bottom) K^+K^-.

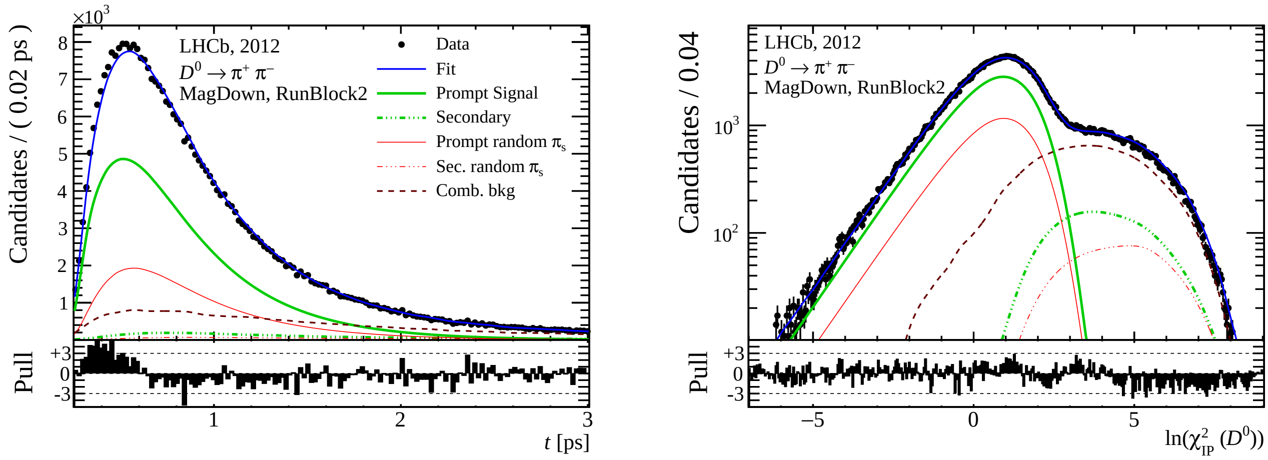


Figure 10.3: Fit projections of (left) D^0 decay-time and (right) $\ln(\chi^2_{\text{IP}})$ for the $\pi^+\pi^-$ final state in the magnet polarity down, run block two and D^0 flavour sub-dataset.

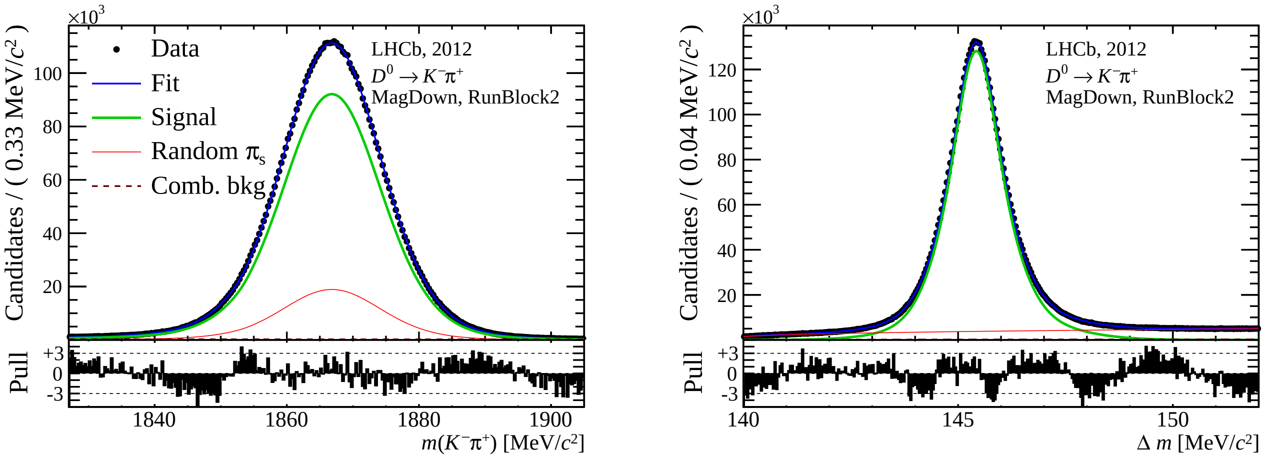


Figure 10.4: Fit projections of (left) $m(D^0)$ and (right) Δm for the $K^-\pi^+$ final state in the magnet polarity down, run block 2 and D^0 flavour sub-dataset.

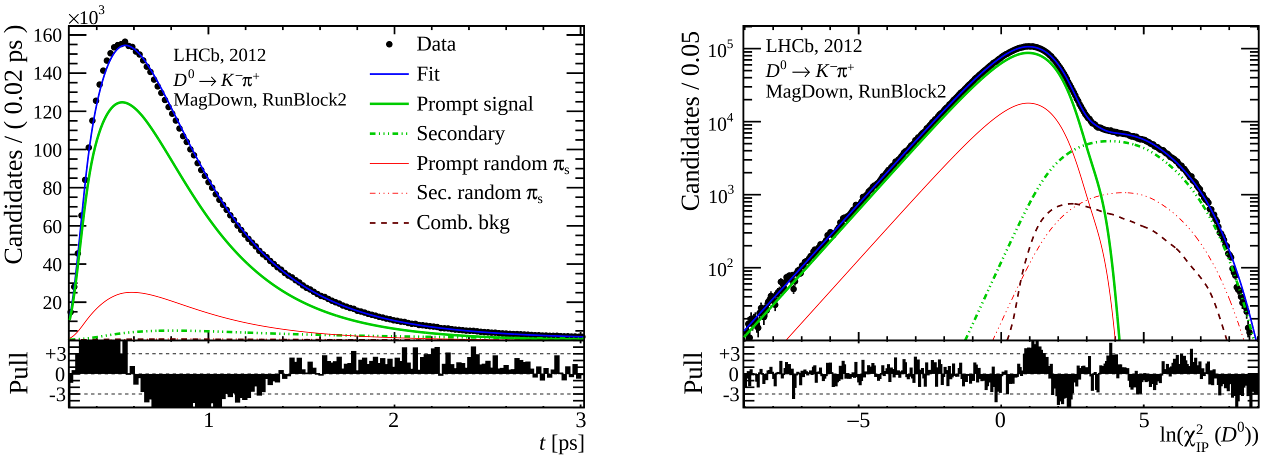


Figure 10.5: Fit projections of (left) D^0 decay time and (right) $\ln(\chi^2_{\rm IP})$ for the $K^-\pi^+$ final state in the magnet polarity down, run block 2 and D^0 flavour sub-dataset.

Table 10.1: Measured values of $\tau(D^0)$ and $\tau(\overline{D}^0)$, split by magnet polarity and run block, for the K^+K^- final state. The averages are weighted by the sample sizes.

	MagDown		
$\tau(K^+K^-)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	410.42 ± 0.77	413.25 ± 0.66	412.20 ± 0.58
$\overline{D}^0$	410.94 ± 0.78	413.54 ± 0.66	411.81 ± 0.58
Average	410.68 ± 0.55	413.40 ± 0.47	412.00 ± 0.41
	MagUp		
$\tau(K^+K^-)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	410.38 ± 0.72	412.05 ± 0.57	413.82 ± 0.70
$\overline{D}^0$	411.01 ± 0.71	412.27 ± 0.55	412.55 ± 0.69
Average	410.70 ± 0.50	412.16 ± 0.40	413.17 ± 0.49

Table 10.2: Measured values of $\tau(D^0)$ and $\tau(\overline{D}^0)$, split by magnet polarity and run block, for the $\pi^+\pi^-$ final state. The averages are weighted by the sample sizes.

	MagDown		
$\tau(\pi^+\pi^-)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	413.87 ± 1.33	414.14 ± 1.11	417.58 ± 1.02
$\overline{D}^0$	412.73 ± 1.33	414.56 ± 1.11	414.02 ± 1.01
Average	413.30 ± 0.94	414.35 ± 0.79	415.80 ± 0.72
	MagUp		
$\tau(\pi^+\pi^-)$ [fs]	RunBlock1	RunBlock2	RunBlock3
D^0	415.16 ± 1.30	413.23 ± 0.99	414.97 ± 1.24
$\overline{D}^0$	415.54 ± 1.29	416.04 ± 0.98	416.15 ± 1.23
Average	415.36 ± 0.92	414.66 ± 0.70	415.57 ± 0.87

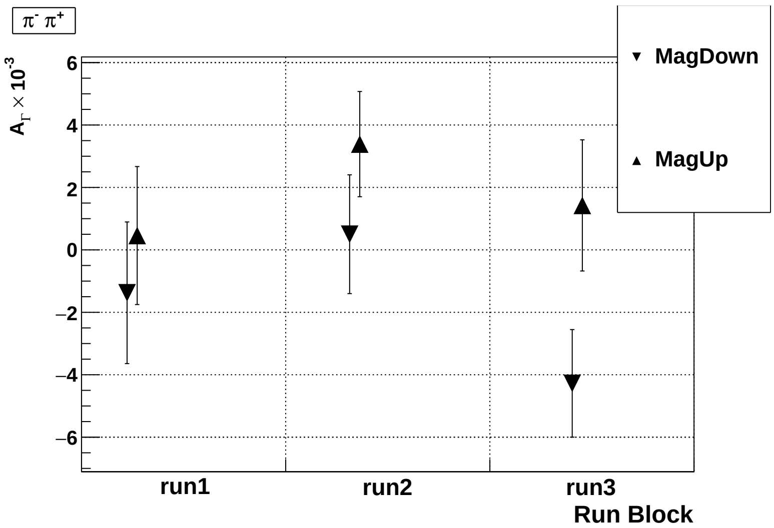


Figure 10.6: Measured values of $A_\Gamma(\pi^+\pi^-)$, split by magnet polarity and run block.

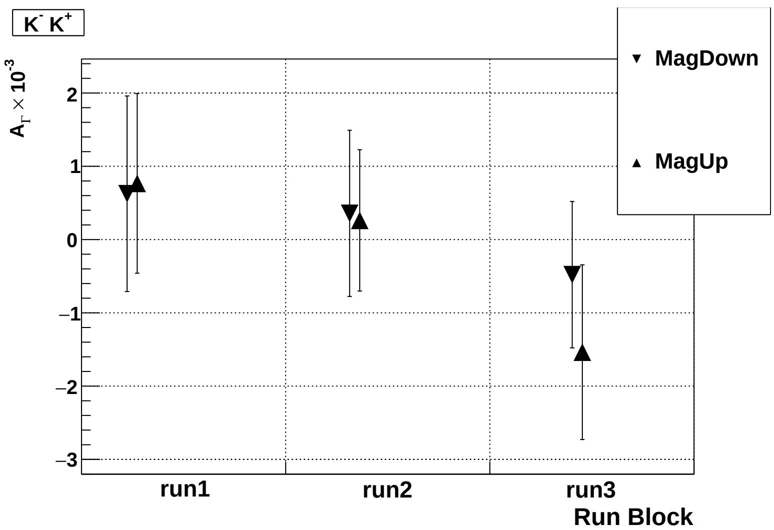


Figure 10.7: Measured values of $A_\Gamma(K^+K^-)$, split by magnet polarity and run block.

10.2.1 Combination

The results from this analysis are combined with those from the analysis of 2011 data [65] to obtain a result for the full LHCb Run 1 prompt data. Both analyses measured the D^0 and $\overline{D}^0$ lifetimes in separate sub-datasets, using the same analysis method. Therefore, a reanalysis of the 2011 data is not deemed necessary and a statistical combination is made. The results from the analysis of 2010 data [64] were initially included, but are of such limited precision that they did not change the overall result.

The measurements are combined by calculating the best linear unbiased estimator [83]. The combinations are performed using the four input measurements, one for each final state from each year of data taking. Their statistical uncertainties are assumed to be uncorrelated, while the systematic uncertainties are considered to be fully correlated. The actual correlation may not be 100% in all cases, but a high correlation is certain given that the analysis method and underlying sources of systematic uncertainty are shared. Systematic uncertainties assigned for the correlations between variables and the low decay-time mismodelling, presented in Sections 9.4.3 and 9.4.4, were not assessed in the 2011 analysis. These systematic uncertainty values from 2012 are assigned directly to the 2011 measurements. Overall, the relatively small impact of the systematic uncertainties on the final result does not merit a more detailed treatment. The results are shown in Table 10.3.

The averaged values for both final states are,

$$\begin{aligned} A_\Gamma(K^+K^-) &= (-0.14 \pm 0.37 \pm 0.10) \times 10^{-3}, \\ A_\Gamma(\pi^+\pi^-) &= (0.14 \pm 0.63 \pm 0.15) \times 10^{-3}. \end{aligned}$$

Combining both final states yields,

$$A_\Gamma = (-0.07 \pm 0.32 \pm 0.11) \times 10^{-3}.$$

The difference between the final states is given by,

$$\Delta A_\Gamma = (0.28 \pm 0.73 \pm 0.18) \times 10^{-3},$$

where in this case the conservative assumption of no correlation between the systematic uncertainties is assumed. All of these measurements are consistent with zero, and thus show no evidence of *CP* violation.

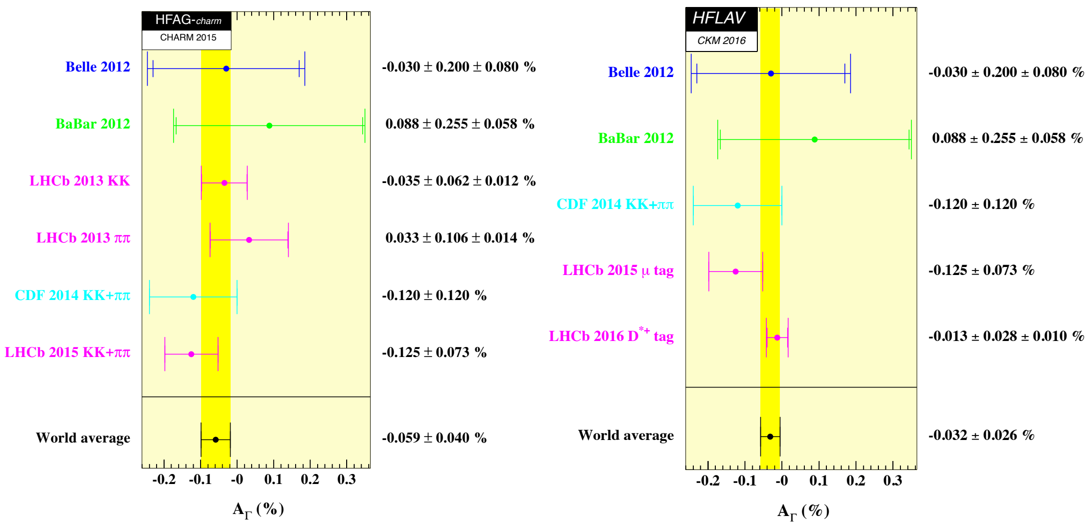


Figure 10.8: World average measurement of A_Γ, compiled by the heavy flavour averaging group (HFLAV, formerly HFAG), from (left) the CHARM 2015 conference, before the inclusion of the binned measurement of A_Γ and (right) the world average from the CKM 2016 conference, which includes the binned measurement [6].

10.3 Conclusion

The measurement described here is one of the two best measurements of A_Γ that has been made to-date. The other measurement, detailed in [66,67], uses the same LHCb Run 1 dataset as is used in this work, but employs a binned fit method. The method presented in this thesis is known as the "unbinned" method, while the complimentary approach is the "binned" method. In the complimentary method, the number of signal events is determined from fits to Δm in bins of decay-time. The asymmetry of the number of D^0 and $\overline{D}^0$ events is calculated in each decay-time bin. The slope of a linear fit to the asymmetry against decay-time is then equal to $-A_\Gamma$. Acceptance effects, namely the charge detection asymmetry of the soft pion, are removed with a reweighting procedure. This affects the intercept of the linear fit to the asymmetry, but not the value of A_Γ. In the binned method, a $\ln(\chi^2_{\text{IP}})$ cut is applied at 2 to remove the majority of the secondary events. A systematic is assigned to account for residual secondary contamination. The resulting values of A_Γ for the two final states and the total average are,

$$\begin{aligned} A_\Gamma(K^+K^-) &= (-0.30 \pm 0.32 \pm 0.10) \times 10^{-3}, \\ A_\Gamma(\pi^+\pi^-) &= (\ \ 0.46 \pm 0.58 \pm 0.12) \times 10^{-3}, \\ A_\Gamma &= (-0.13 \pm 0.28 \pm 0.10) \times 10^{-3}. \end{aligned} \tag{10.1}$$

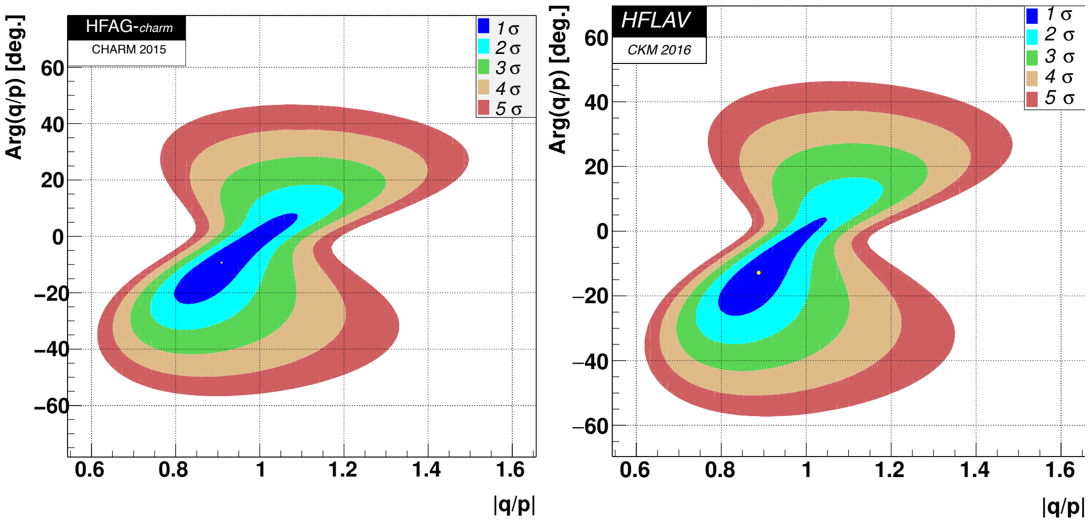


Figure 10.9: Fit of $|q/p|$ and $\phi := \arg(q/p)$ using various measurements of *CP* observables, compiled by the heavy flavour averaging group (HFLAV, formerly HFAG), from (left) the CHARM 2015 conference, before the inclusion of the binned measurement of A_Γ and (right) the fit from the CKM 2016 conference, which includes the binned measurement of A_Γ along with new measurements of other observables [6].

Before unblinding either the binned or unbinned measurements, a data-driven study was carried out to determine how correlated the results are. Using the correlation, the difference between the two measurements was calculated to show that both are in agreement. These tests are sufficient for roughly checking the agreement of the two measurements, but are not sufficient for averaging the two results. The gain in sensitivity from averaging the results is likely to be very small and was therefore not considered to be beneficial.

The binned method is significantly simpler than the method presented in this thesis, and yet produces more accurate measurements of A_Γ. This is due in part to the large number of nuisance parameters in the unbinned method fits and the relative complexity of including the swimming acceptance into the fit PDFs. The reweighting procedure of the binned method, along with the ratio inherent to calculating an asymmetry, are sufficient to remove acceptance effects. Similarly, assigning a systematic for residual secondary contamination instead of attempting to remove it with a fit, is sufficient for calculating A_Γ.

The world average measurements of A_Γ before and after including the binned measurement are shown in Fig. 10.8. It can be seen that the LHCb measurements have the biggest effect on the world average. The most recent measurement with the binned method supercedes some of the previous measurements and dominates the world average.

As seen in Eq. (2.52), the values of $|q/p|$ and ϕ depend on A_Γ. Shown in Fig. 10.9 is the contour plot of $|q/p|$ against ϕ resulting from a fit to various experimental measurements [6], before and after the inclusion of the binned measurement of A_Γ, along with new measurements of other observables. The A_Γ measurement does not have a large impact on the fit relative to other measurements, but it does contribute to constraining $|q/p|$.

The binned method produces more accurate measurements of A_Γ, with a much simpler analysis. However, it is unsuitable for measuring another important CP observable from this dataset, denoted y_{CP} as in Eq. (2.47). This is due to the different acceptance effects in the three relevant decay channels. The K^+K^-, $\pi^+\pi^-$ and $K^-\pi^+$ acceptances are not necessarily the same, thus the acceptance effects are not guranteed to cancel in y_{CP} as they do in A_Γ. In theory the unbinned method can deal with these acceptance differences; indeed, y_{CP} has already been measured with 2010 data using this method [64]. Further measurements of y_{CP} with LHCb data are underway; however, it is not yet clear from these studies if the unbinned method can suficiently account for the different final state acceptances. In Part III, new methods for directly measuring the D^0 lifetime with a binned approach are studied. The different acceptances of the final states will no longer be an issue for y_{CP} measurements if LHCb is able to directly measure lifetimes.

Table 10.3: Combination of A_Γ measurements from 2011 and 2012 data and the decays $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$.

Syst. Uncert. ($\times 10^{-3}$)	Input				Combinations		
Year	11	11	12	12	$11+12$	$11+12$	$11+12$
Decay	K^+K^-	$\pi^+\pi^-$	K^+K^-	$\pi^+\pi^-$	K^+K^-	$\pi^+\pi^-$	$K^+K^- + \pi^+\pi^-$
A_Γ ($\times 10^{-3}$)	−0.35	0.33	−0.03	0.03	−0.14	0.14	−0.07
correlations	0.02	0.07	0.02	0.07	0.02	0.07	0.03
specific backgrounds	0.02	0.00	0.01	0.00	0.01	0.00	0.01
secondary contamination	0.07	0.07	0.08	0.08	0.08	0.08	0.08
combinatoric background	0.02	0.04	0.02	0.10	0.02	0.08	0.03
turning point accuracy	0.09	0.11	0.03	0.02	0.05	0.05	0.05
random soft pion background	0.00	0.00	0.02	0.03	0.01	0.02	0.01
decay-time resolution	0.00	0.00	0.01	0.01	0.01	0.01	0.01
low decay-time mismodelling	0.03	0.05	0.03	0.05	0.03	0.05	0.03
total systematic	0.12	0.16	0.10	0.16	0.10	0.15	0.11
total statistical	0.62	1.06	0.46	0.79	0.37	0.63	0.32
Total	0.63	1.07	0.47	0.81	0.38	0.65	0.34

Part III

New methods for charm lifetime measurements

Chapter 11

Overview

The method discussed in Part II for measuring charm lifetimes has a number of impediments. The swimming algorithm used to remove lifetime biases in the event selection is CPU intensive. As larger datasets are collected at LHCb it will no longer be feasible to store and process each event with the swimming algorithm. It is also clear that this method doesn't fully account for one or both of the decay-time acceptance and the secondary contamination, as shown by the biased measurement of the D^0 lifetime in Part II. This is not a problem for the A_Γ measurement, as it is known that the bias cancels in the ratio of lifetimes, but it does pose a problem for other CP violation measurements, such as y_{CP}. This observable requires the calculation of the ratio of lifetimes in different final states, as discussed in Section 2.5, where the acceptances and backgrounds differ.

The complex fit procedure for these measurements requires the dedicated efforts of a number of analysts for long periods of time. During this time LHCb continues to collect more data. It is now the case that half of the LHCb dataset has not yet been analysed, as so much time was spent analysing the first half. The accuracy of many charm CP violation measurements are limited by statistical uncertainties, thus the accuracy can be significantly increased by considering all available data. Simplified analysis methods are required such that analysis of the data can keep pace with the rate of data collection. Presented here are potential methods for simplifying measurements of charm meson lifetimes using new triggers that were specifically developed to reduce decay-time biases. As a demonstration, a blinded measurement of the D^0 meson lifetime is presented that is the second most sensitive determination to-date.

The Cabibbo favoured (CF) decay mode $D^0 \to K^-\pi^+$ is used to measure the D^0 meson lifetime, where charge conjugation is assumed throughout. Its

Table 11.1: Results for $\tau(D^0)$ that contribute to the world average value. The errors quoted are statistical and systematic uncertainties, respectively.

Collaboration	Year	Num. D^0 Candidates	$\tau(D^0)$ [fs]
FOCUS [84]	2002	210,000	$409.6 \pm 1.1 \pm 1.5$
SELEX [85]	2001	10,000	$407.9 \pm 6.0 \pm 4.3$
E791 [86]	1999	35,000	$413 \pm 3 \pm 4$
CLEO [87]	1999	25,000	$408.5 \pm 4.1^{+3.5}_{-3.4}$
E687 [88]	1994	16,000	$413 \pm 4 \pm 3$

high branching fraction of $(3.89 \pm 0.04)\%$ [9], results in ample measurement statistics. A dataset collected by LHCb in 2016, with an integrated luminosity of $1.6\,\mathrm{fb}^{-1}$, is used. The D^0 production decay $D^{*+} \to D^0\pi^+$ is used to give separation between prompt and secondary D^0 mesons, where secondary D^0 mesons are those produced in a B meson or other long-lived particle decay. Such events result in misreconstructed flight distances as the candidate D^0 mesons are assumed to come directly from the primary interaction vertex (PV), and thus they have wrongly measured decay times.

The first and main method presented in Chapters 12 to 14 removes events with a requirement of $\ln(\chi^2_{\mathrm{IP}}) > 1.6$ to separate prompt D^0 decays from secondary decays. This is the same strategy used by the binned A_Γ measurement, first discussed in Section 10.3. The data is split into eight bins of measured D^0 decay time in the range $[0.4, 1.2]\,\mathrm{ps}$. The number of prompt signal events are extracted from each decay-time bin via a binned maximum likelihood fit to the mass difference $\Delta m = m(D^*) - m(D^0)$ after tight selection criteria are applied to the D^0 meson mass, $m(D^0)$. The final analysis step is a fit to the yield ratio $N_{\mathrm{obs},i}/N_{\mathrm{exp},i}$, where the numerator is the number of observed prompt signal events in each decay-time bin i, and the denominator is the number of expected events in each bin, given the PDG lifetime of $(410.1 \pm 1.5)\,\mathrm{fs}$ [9]. An alternate method for separating prompt and secondary decays uses machine learning algorithms instead of a $\ln(\chi^2_{\mathrm{IP}})$ selection requirement; this method is presented in Chapter 15.

The current world average measurement of the D^0 lifetime is dominated by the FOCUS collaboration measurement [84]. Charm mesons were produced in the experiment with a 180 GeV photon beam incident on a beryllium oxide target. Candidate particles were found by their characteristic secondary vertices and the quantity f/σ_f, where f is the flight distance, is used to separate prompt and secondary decays. A binned maximum likelihood fit to the $K^-\pi^+$ invariant mass gives the prompt signal yields and a binned fit to these yields against decay time returns the lifetime. The results of the FOCUS measurement along with other measurements that contribute to the world average value are shown in Table 11.1. The LHCb dataset used for this measurement contains about 14 million events such that the statistical uncertainty is significantly smaller than for previous results. An LHCb measurement of $\tau(D^0)$ has some disadvantages when comparing to previous measurements, namely the challenging high energy collision environment. This results in higher systematic uncertainties and a similar overall accuracy.

Chapter 12

Datasets and selection

The D^0 candidate selection for this analysis is presented here, along with the new Run 2 trigger lines that attempt to reduce decay-time biases. The backgrounds contaminating the prompt signal decays are the same as those described for the measurement of A_Γ in Chapter 4. However, no distinction is made in this analysis between the combinatorial background and the random soft pion background. They are collectively known as the "remaining" backgrounds throughout.

12.1 Decay-time unbiased selections

This section describes the online selections applied to the data to produce a dataset of $D^0 \to K^-\pi^+$ decays with minimal decay-time biases. Recall that the trigger strategy of LHCb is considerably different in Run 2 compared to Run 1, as was discussed in Section 3.2.4. The offline selection known as the "stripping" is not performed for this analysis and the calibration, alignment, and PID information are available to HLT2 triggers.

It was learnt from Run 1 analyses of $D^0 \to K^-\pi^+$ and similar decays that the $\chi^2_{\rm IP}$ requirement used to reduce short-lived background near the PV resulted in non-trivial decay-time biases, as was seen in Part II. These requirements were deemed to be necessary to reduce short-lived backgrounds to an acceptable level in the trigger. Without such selection requirements the number of events passing the trigger would have been unmanageable. A direct selection requirement on decay-time in the trigger, which would trivially bias the decay-time, was not possible in Run 1 as the online and offline measurements of the decay-time could be significantly different. This difference in Run 1 is a consequence of different alignments and calibrations being used in the online and the offline reconstruction

Table 12.1: Decay-time unbiased HLT1 trigger selection criteria.

Observable	Selection requirement
Both D^0 daughter p_T	$> 600\,\mathrm{MeV}/c$
Both D^0 daughter momentum	$> 4000\,\mathrm{MeV}/c$
Both D^0 daughter track $\chi^2_{/\mathrm{ndf}}$	< 2
Largest D^0 daughter p_T	$> 900\,\mathrm{MeV}/c$
D^0 daughters' DOCA	$< 0.1\,\mathrm{mm}$
D^0 mass range	$[1804, 1924]\,\mathrm{MeV}/c^2$
D^0 p_T	$> 1800\,\mathrm{MeV}/c$
D^0 vertex fit $\chi^2_{/\mathrm{ndf}}$	< 10
D^0 decay time	$> 0.25\,\mathrm{ps}$
D^0 PV DIRA	> 0.99

algorithms. Due to the restructuring of the trigger in Run 2, and improvements in reconstruction and measurement algorithms, the best measurements of the decay-time are available to the HLT trigger. As a result, a simple selection requirement on the decay-time can be made at the trigger stage to remove short-lived background instead of relying on χ^2_{IP} requirements.

A series of special triggers were designed for Run 2 that produce samples of various decays with minimal decay-time biases, known simply as "lifetime unbiased triggers". Each one consists of a HLT1 trigger and a corresponding HLT2 trigger. The relevant HLT1 trigger for this analysis is, `Hlt1CalibTrackingKPi`, which is also described in reference [89]. As can be seen in Table 12.1, the selection requirements applied in this new HLT1 trigger are more extensive than those made in the Run 1 equivalent trigger of Table 5.1.

At HLT2 the candidates are required to be TOS with respect to the `Hlt2CharmHadDstp2D0Pip_D02KmPip_LTUNBTurbo` line. The selection and reconstruction criteria for this trigger are more similar to the full offline reconstruction of Run 1 than to the corresponding HLT2 selection requirements. Using this new trigger format the RICH based PID selection is now available at the trigger level in Run 2. The new HLT2 trigger selection requirements are shown in Table 12.2, and can be contrasted to the Run 1 selection requirements of Tables 5.2 and 5.3. Note also that the DIRA requirement with respect to the PV is loosened in the new triggers, which reduces the low decay-time acceptance effect described in Section 5.4.

Table 12.2: Decay-time unbiased HLT2 trigger selection criteria.

Observable	Selection requirement
Both D^0 daughter p_{T}	$> 800\,\mathrm{MeV}/c$
Both D^0 daughter momentum	$> 1000\,\mathrm{MeV}/c$
Largest D^0 daughter p_{T}	$> 1200\,\mathrm{MeV}/c$
Both D^0 daughter track $\chi^2_{/\mathrm{ndf}}$	< 3.0
D^0 daughters' DOCA	$< 0.1\,\mathrm{mm}$
Kaon PIDK	> 10
Pion PIDK	< 0
D^0 p_{T}	$> 2.0\,\mathrm{GeV}/c$
D^0 vertex fit $\chi^2_{/\mathrm{ndf}}$	< 10.0
D^0 PV flight distance $\chi^2_{/\mathrm{ndf}}$	> 40.0
D^0 PV DIRA	> 0.99
D^0 mass	$[1774, 1954]\,\mathrm{MeV}/c^2$
D^0 decay-time	$> 0.25\,\mathrm{ps}$
Δm	$[130, 160]\,\mathrm{MeV}/c^2$
D^{*+} vertex fit $\chi^2_{/\mathrm{ndf}}$	< 25
π_s vertex fit $\chi^2_{/\mathrm{ndf}}$	< 3
π_s p_{T}	$> 100\,\mathrm{MeV}/c$
π_s momentum	$> 1\,\mathrm{GeV}/c$

12.2 Offline selection

The offline requirements applied to the data are described here. Tighter selection requirements than in the HLT are made around the $m(D^0)$ and Δm peaks at $[1835, 1890]\,\mathrm{MeV}/c^2$ and $[139.5, 160]\,\mathrm{MeV}/c^2$, respectively, to reduce the combinatorial and random soft pion sidebands. Tighter constraints are made on the quality of the D^{*+} decay vertex by removing events with $\chi^2{}_{/\mathrm{ndf}} > 15$ and $\sigma_z > 20\,\mathrm{mm}$.

A requirement of a decay radius of less than 4 mm is applied to the D^0 candidate in order to reduce the level of candidates produced by material interactions in the RF foil. The decay radius is defined as the radial distance between the D^0 vertex and the associated PV.

Only events with one candidate PV are used in the analysis. This is to remove events where the decay time is misreconstructed due to the candidate being associated to the wrong PV. The number of PVs condition and the radial flight distance requirement can create upper decay-time acceptance effects. A long-lived D^0 is more likely to be associated with the wrong vertex as it travels further from the correct vertex. Similarly, the geometry of the VELO and track finding algorithms can cause upper decay-time effects. As a result the candidate decay-time range is restricted to be less than 1.2 ps, which precludes these effects. A lower decay-time requirement of $t > 0.4\,\mathrm{ps}$ is applied to remove any potential low decay-time biases.

12.3 Datasets and binning

The dataset has an integrated luminosity of $1.6\,\mathrm{fb}^{-1}$ and was collected by LHCb in 2016. The polarity of the LHCb magnet during data taking is used to define two sub-datasets, polarity up and polarity down. Before applying selection requirements both the magnet polarity up and down samples contain about 50 million candidates. Table 12.3 shows the number of events in each of these sub-datasets after offline selection.

The data is split into eight decay-time bins with their widths set according to the PDG world average lifetime. The width of the first bin is fixed to be 50 fs, which is the decay-time resolution of LHCb, as was used in the analysis of Part II (see Section 8.6). The width of the subsequent bins are chosen such that the integral of an exponential decay with the PDG lifetime has the same value in each bin. The last bin is defined by the maximum of the decay-time range and does not have the same integral value as the other bins. As a consequence of this scheme, if the real measured lifetime is equal to the PDG value then there will be an equal number of prompt signal events in each bin, except the last bin. Otherwise, the prompt signal data would have an approximate linear distribution where the slope is related to the lifetime.

Table 12.3: Number of candidates (n_{cand}) present for the magnet polarity up and down sub-datasets after the offline selection has been applied but before the decay-time range selection is applied. Also shown are the number of candidates after all online and offline selection requirements have been applied.

n_{cand} (Millions)	Before decay-time range selection	After all selections
polarity up	12.5	6.8
polarity down	13.0	7.1

Due to a number of problems with LHCb Run 2 simulation, there are no suitable full GEANT4 based simulation samples available for this analysis. Samples generated under Run 1 conditions are used instead. These samples are made with the beam conditions of 2012 data taking and the same reconstruction and selection as in Part II. Specifically, the decay-time biasing selections of the trigger and stripping are applied to these samples, resulting in several problems for the studies to follow. Two distinct samples are used, one of true prompt events and one of true secondary events. The prompt sample contains about 800k events while the secondary sample contains 150k events.

12.4 Secondary background reduction

The method presented here is similar to the binned A_Γ measurement discussed in Section 10.3. A selection requirement is applied to $\ln(\chi^2_{\text{IP}})$ to remove the majority of secondary contamination and a systematic uncertainty is later assigned to account for any residual contamination. The selection requirement value is chosen by comparing Run 1 prompt and secondary simulation samples and choosing a requirement that optimises the metric $S/\sqrt{B}$ where S is the number of prompt signal events and B is the number of secondary background events in the simulation

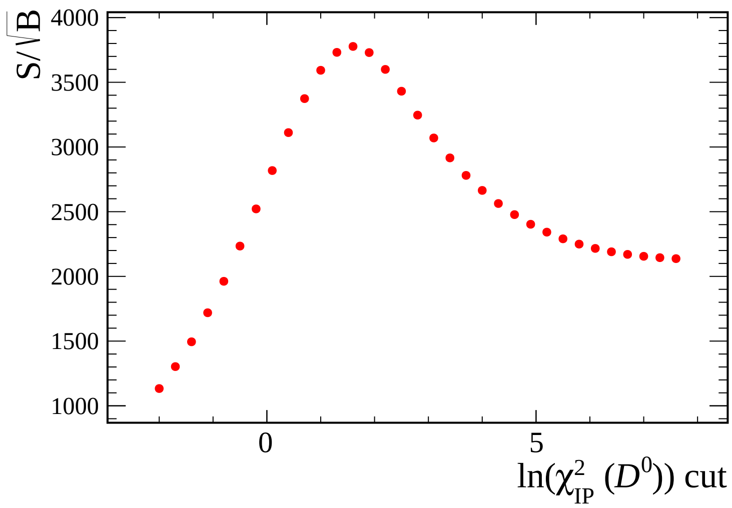


Figure 12.1: Distribution of the separation metric $S/\sqrt{B}$ for prompt and secondary simulated events against $\ln(\chi^2_{\text{IP}})$ and assuming that secondary events are 10% of the total number of signal events. The optimum occurs at a $\ln(\chi^2_{\text{IP}})$ selection requirement of < 1.6.

samples. The percentage of secondary decays present is not known, but must be estimated from the data. The percentage of secondary events with respect to the total number of signal events is assumed to be 10%, which is similar to the 7% value observed in the analysis of Part II. The value is motivated by comparing the distributions of $\ln(\chi^2_{\mathrm{IP}})$ from the real data and the simulation samples then scaling the relative fraction of the secondary sample until the distributions match. The separation metric is shown in Fig. 12.1 against $\ln(\chi^2_{\mathrm{IP}})$, where the optimum selection requirement is found to be < 1.6. Table 12.3 also shows the number of events in the dataset after removing events with $\ln(\chi^2_{\mathrm{IP}}) > 1.6$.

Chapter 13

Fit method

The next step of this analysis is to extract the number of prompt signal events in bins of decay-time and fit to these decay-time bins to determine the D^0 lifetime. The Δm observable is used to separate the prompt signal from the remaining backgrounds. The fits to Δm in bins of decay-time are detailed in Section 13.1, while the fit to the number of prompt signal events across these bins is discussed in Section 13.2.

13.1 Fits to Δm

To extract the number of prompt signal events and remove the combinatorial and random soft pion backgrounds, unbinned maximum likelihood fits are performed to Δm in each decay-time bin. The prompt signal component is modelled as a Johnson SU along with two Gaussians. The Johnson SU is given by,

$$f(x) = \frac{\delta}{\sigma\sqrt{2\pi}\sqrt{z^2+1}} \exp\left\{-\frac{1}{2}\left(\nu + \delta \ln\left(z + \sqrt{z^2+1}\right)\right)^2\right\}, \tag{13.1}$$

where $z = \frac{x-\mu}{\sigma}$, μ is the most probable value, σ characterises the width of the distribution and ν and δ influence the shape. The shape parameters are highly correlated and are thus not independent. To solve this, the ν parameter is fixed in the fits to a value of 0.38 that was obtained from preliminary fits. All three signal components share the same mean but have different widths. There are two fractions that determine the relative size of each of the three components. The fraction that determines the size of the second Gaussian relative to the other Gaussian and the Johnson SU is fixed to 0.05 from preliminary fits. Thus there are six free parameters in the description of the prompt signal component.

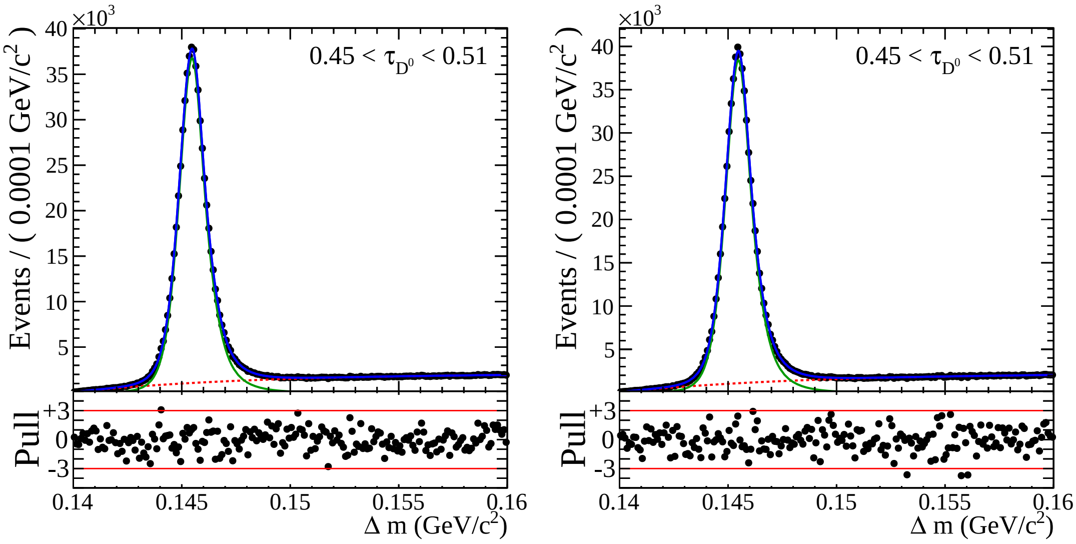


Figure 13.1: Fits to the Δm distributions of (left) magnet polarity up and (right) magnet polarity down, to extract the number of prompt signal events. The fit components are (blue) the total fit, (green) the prompt signal component and (dashed red) the remaining backgrounds. The bottom frame in each plot shows the pull distribution, or the difference between the total fit and the data, divided by the data uncertainty.

The remaining background component, which contains randomly assigned soft pions and combinatorial background, is described with a single root exponential PDF, as given by Eq. (7.4). The decay parameter cannot be well constrained so is fixed to a large negative value such that it has no effect on the fit. The threshold parameter is free to float near the value of the pion mass and the power parameter is also free to float, thus giving two free parameters.

The Δm fits for the second decay-time bin are shown in Fig. 13.1. All of the Δm fits in each decay-time bin are shown in Appendix C.

13.2 Lifetime fits

To construct the ratio $N_{\text{obs},i}/N_{\text{exp},i}$, the number of observed prompt events from the Δm fits are divided by the integral of a control exponential, $f_{PDG}(t) = e^{-t/\tau_{\text{PDG}}}$, in each bin. The control exponential distribution is normalised across the entire decay-time range such that it contains the same number of events as the observed data, $\sum_i N_{\text{obs},i}$. Thus the value in each decay-time bin is given by,

$$\frac{N_{\text{obs},i}}{N_{\text{exp},i}} = N_{\text{obs},i} \left[\int_{t_{i,\min}}^{t_{i,\max}} f_{\text{PDG}}(t) \right]^{-1} \left[\sum_i N_{\text{obs},i} \right]^{-1} \left[\int_{t_{\min}}^{t_{\max}} f_{\text{PDG}}(t) \right], \tag{13.2}$$

where $t_{i,\mathrm{min}}$ and $t_{i,\mathrm{max}}$ denote the upper and lower decay-time limits of a bin.

The function used to fit the data is a ratio of the integrals of two exponential functions within each bin,

$$f_{\mathrm{fit},i}(t) = \frac{\int_{t_{i,\mathrm{min}}}^{t_{i,\mathrm{max}}} e^{-t/\tau}}{\int_{t_{i,\mathrm{min}}}^{t_{i,\mathrm{max}}} f_{\mathrm{PDG}}(t)}. \tag{13.3}$$

This function produces a series of steps such that it can be compared to the binned distribution of the data. Once the binning scheme is chosen, the fit function has a single parameter, τ, which is the D^0 meson lifetime.

The data is blinded to avoid analyst bias by using an unknown value of the control lifetime. A new control lifetime is chosen at random from a normal distribution with mean equal to the PDG value and width equal to the PDG uncertainty of 1.5 ps. The same control lifetime is chosen for both magnet polarity up and down sub-datasets such that the two samples are consistent. In this way, the $\tau(D^0)$ value extracted from the fits is within about 2×1.5 ps from its PDG value. The lifetime fits for both magnet polarity samples are shown in the next chapter in Fig. 14.4.

Chapter 14

Results and systematic uncertainties

14.1 Systematic uncertainties and cross-check

Cross-checks of the analysis and tests for systematic biases in the method are presented here. The magnitude of each systematic effect and its source are shown in Table 14.1. Systematic effects are investigated using alternative fit configurations and generated pseudo-experiments known as "toys". In the following sections, the sources of these systematic effects are discussed along with the two tests of each effect that returned the largest systematic biases. For each effect, the larger of the biases is taken as the systematic uncertainty.

The toys are generated with three component classes, prompt signal, secondary background and a third component representing the remaining backgrounds. In the Δm observable, the prompt and secondary events are generated according to the signal PDF from the Δm fits to data, while the third component is generated with the background component from the Δm fits. The fit parameters from one of the eight decay-time bins of the magnet polarity down sample are used to generate the toy data.

The decay-time PDF of prompt signal is a single exponential with the PDG world average lifetime central value of 410.1 fs, while the remaining background is generated with a sum of two exponentials. The first exponential has a lifetime of 370 fs and the second has 150 fs with a relative fraction between these PDFs of 0.3, such that the first is dominant. These parameter values roughly reproduce the decay-time distribution of this background as seen in data.

The distribution of $\ln(\chi^2_{\text{IP}})$ is also generated such that the cut on this observable can be simulated. As only a cut on $\ln(\chi^2_{\text{IP}})$ is used in the analysis, the distributions of the components in $\ln(\chi^2_{\text{IP}})$ need not be modelled accurately in the toy samples. The prompt signal $\ln(\chi^2_{\text{IP}})$ component is generated with a smeared $\ln(\chi^2_{\text{IP}})$ PDF with parameters obtained from a fit to the Run 1 simulation. The secondary component has a conditional 2D PDF in decay-time and $\ln(\chi^2_{\text{IP}})$ to simulate the time dependence of $\ln(\chi^2_{\text{IP}})$. The form of this PDF is obtained from fits to Run 1 simulation in bins of decay-time as described in Appendix A.2. An arctan decay-time dependence is used for the mean of the smeared $\ln(\chi^2_{\text{IP}})$ PDF, while the tail parameters have no decay-time dependence.

The remaining background comprises of three components in real data, combinatorial background, and prompt and secondary random soft pion backgrounds. The combinatorial background from real data has a wide distribution that peaks at about 4 in $\ln(\chi^2_{\text{IP}})$, while the prompt and secondary random soft pion components have the same shapes as prompt and secondary signal, respectively. As a simplification these remaining background components are described in the generated toys as a single Gaussian with mean at one and a width of 1.5, such that the cut at a $\ln(\chi^2_{\text{IP}})$ value of 1.6 removes a reasonable portion of the remaining background.

The secondary decay-time projection is a convolution of two exponentials plus a third exponential, as was seen in the analysis of Part II. The parameters are taken from fits of the secondary decay-time in the Run 1 simulation.

The three components of the toy data are generated such that the prompt accounts for 58.5% of the total generated data, the secondary 6.5% and the remaining background 35%. These percentages result in the signal component (prompt plus secondary) containing 10% secondary contamination as required. In total, 10 million events are generated, which is comparable to the number of candidates in the real data.

A control toy test was done in which only the prompt signal component was generated, to check the consistency of the method. The results of this control test are shown in Table 14.2 where the generated prompt lifetime was set to be $\pm 1\sigma$ and $\pm 2\sigma$ of the PDG world average value. The generated pseudo-data for the five samples of this test are fully correlated. Similarly, the samples shown in toy tests to follow are also fully correlated. No significant bias is seen in the control toy test, indicating that in the absence of backgrounds the method is not biasing the result. As will be seen in the following sections, the addition of backgrounds introduce large biases into the analysis method.

Table 14.1: Summary of systematic uncertainties on the measurement of the D^0 lifetime.

Effect	Systematic bias [fs]
Secondary contamination	± 2.8
Kinematic splits	± 3.0
Remaining backgrounds	± 0.5
Total systematic uncertainty	± 4.1
Total statistical uncertainty	± 0.3

Table 14.2: Results of a series of control toy tests generated with a prompt component only. In this test different prompt signal lifetimes were generated that were $\pm 1\sigma$ or $\pm 2\sigma$ from the PDG world average value of the D^0 lifetime. The first and second columns give the generated and measured lifetime, respectively. The third column shows the lifetime fit χ^2 and the final column shows the bias of the measured value with respect to the generated value.

Generated $\tau(D^0)$ [fs]	measured $\tau(D^0)$ [fs]	fit $\chi^2_{/\mathrm{ndf}}$	bias [fs]
407.10	407.13 ± 0.34	8.5/7	0.03
408.60	408.64 ± 0.34	8.2/7	0.04
410.10	410.09 ± 0.34	6.8/7	-0.01
411.60	411.65 ± 0.35	7.4/7	0.05
413.10	413.12 ± 0.35	7.5/7	0.02

14.1.1 Secondary contamination

Contamination of the prompt signal events by secondary D^0 decays gives a large systematic bias to the analysis method, as was expected. This irreducible background, which has a tail under the prompt signal peak in $\ln(\chi^2_{\mathrm{IP}})$, is always challenging to account for.

The largest systematic effects come from fitting toy datasets generated with prompt and secondary components, as shown in Table 14.3. A bias is observed at a level of 2.8 fs, with the full value assigned as a systematic uncertainty.

The variation of the bias with the generated D^0 lifetime, in a range of $\pm 2\sigma$ around the PDG value, was tested. No strong dependence on the D^0 lifetime was

Table 14.3: Results of a series of toy tests generated with prompt and secondary components only. In this test different prompt signal lifetimes were generated that were $\pm 1\sigma$ or $\pm 2\sigma$ from the PDG world average value of the D^0 lifetime. The first and second columns give the generated and measured lifetime, respectively. The third column shows the lifetime fit χ^2 and the final column shows the bias of the measured value with respect to the generated value.

Generated $\tau(D^0)$ [fs]	measured $\tau(D^0)$ [fs]	fit $\chi^2_{/\mathrm{ndf}}$	bias [fs]
407.1	409.92 ± 0.36	12.9/7	2.82
408.6	411.39 ± 0.36	13.0/7	2.79
410.1	412.80 ± 0.36	12.2/7	2.70
411.6	414.35 ± 0.36	13.1/7	2.75
413.1	415.78 ± 0.36	13.2/7	2.68

observed, as also shown by Table 14.3.

These fits have worse $\chi^2_{/\mathrm{ndf}}$ values than the control test results, which indicate that the distributions are less likely to be linear[1]. Such non-linear distributions are characteristic of secondary contamination or acceptance effects.

Another test that returns the second largest systematic uncertainty is a series of toys generated with prompt and secondary events only, where the secondary fraction was varied. As expected, the bias to the measured lifetime increases with increasing secondary fraction. The largest generated fraction was 0.10, resulting in a bias of 2.70 fs.

A number of other tests were done that returned smaller systematic biases than those described above. One such test was done with a simplified generated secondary component in the toy data. A single exponential was used instead of the combination of three exponentials. The lifetime parameter of the single exponential, the fraction of the secondary component, and the prompt lifetime, were all varied in a series of tests and all biases were found to be smaller than those outlined above.

[1] The true distribution is a ratio of two exponentials, which is approximately linear in the binning scheme used and thus referred to as "linear" throughout.

14.1.2 Kinematic splits

To test for residual acceptance effects and background contamination, the datasets are split into regions of D^0 momentum, D^0 transverse momentum, and pseudorapidity. The data are split by D^0 momentum, transverse momentum p_T and pseudorapidity, η. It is found that the measured lifetime depends strongly on these observables as shown in Figs. 14.1 to 14.3 for momentum, transverse momentum and pseudorapidity, respectively. Selection criteria of $m(D^0) \in [1840, 1890]\,\mathrm{MeV}/c^2$ and $\Delta m \in [143, 148]\,\mathrm{MeV}/c^2$ are applied to the data to separate signal from combinatorial background and random soft pion background events. Similarly a cut at $\ln(\chi^2_{\mathrm{IP}}) < 1$ is used to select a sample with an enhanced fraction of prompt signal and $\ln(\chi^2_{\mathrm{IP}}) > 2$ is used to select a sample containing more secondary background. There is contamination between the three samples, but they should be sufficient for illustrating correlations. A linear fit to the profile of momentum versus decay-time for prompt signal shows there is a significant negative slope, indicating a small but significant dependence. For the other two components, clear non-linear dependencies are evident in the figures. Similarly, the profile of the decay-time against η has a small linear dependence for prompt signal and significant non-linearities for the two background classes. The dependence on p_T is concerning as all three components have significant non-linear dependencies.

The effect of these correlations on the measured lifetime is assessed by splitting the data into bins of each of the three variables and running the analysis on the sub-datasets. The data is split into three momentum bins defined by $[0, 37, 53, 1000]\,\mathrm{GeV}/c$, such that each bin contains approximately the same number of events. Table 14.4 shows the results of this test for the magnet up and the magnet down sub-datasets and for all three momentum bins. A significantly longer lifetime is measured in the first momentum bin.

The corresponding results from the p_T split are shown in Table 14.5 where the most significant bias is due to the first bin and has a value of 3.0 fs. The missing value in the last p_T bin is due to a poor lifetime fit $\chi^2_{/\mathrm{ndf}}$ value as the Δm fit of the last decay-time bin did not converge. The negative bias of the last bin corresponds to the lower average lifetime of the prompt component in high p_T regions. The large positive bias in the low p_T bin likely indicates a contamination of secondary decays, which have a large positive decay-time bias in the low p_T region.

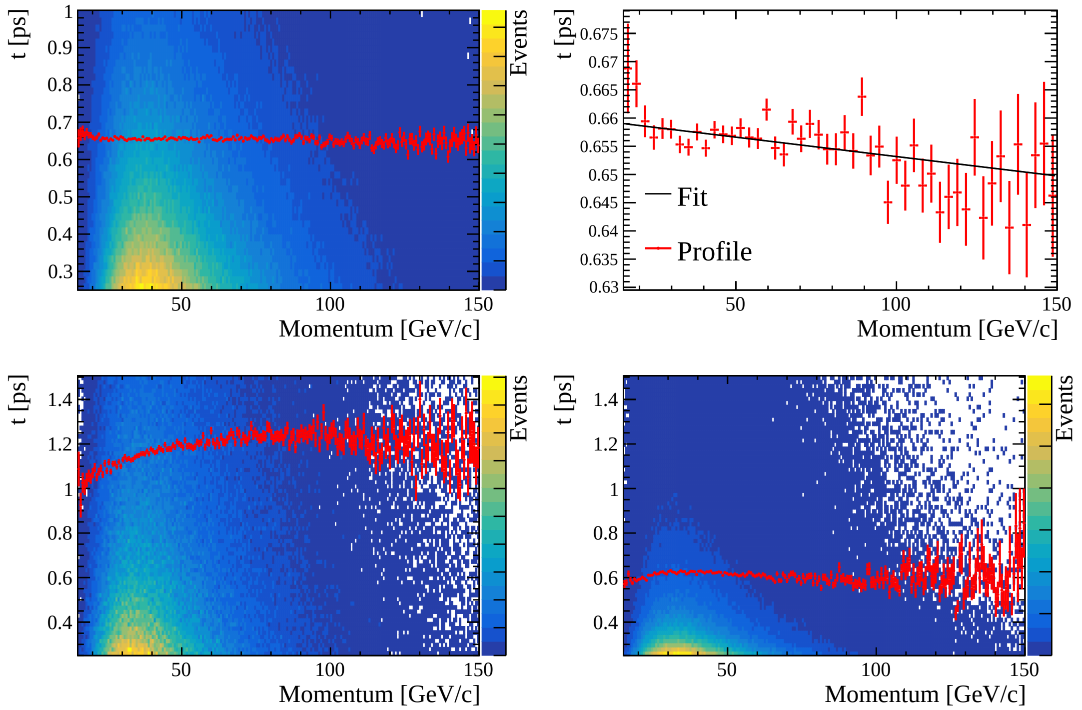


Figure 14.1: Data distributions of momentum against decay-time overlaid with (red points) the profile of decay-time in bins of momentum. All four plots contain data with various selection requirements applied to $m(D^0)$, Δm and $\ln(\chi^2_{\rm IP})$, to select enriched samples of (top left and top right) prompt signal, (bottom left) secondary background and (bottom right) remaining backgrounds. A linear fit to the prompt sample profile (top right) shows that the prompt sample has a small but significant momentum dependence. The two backgrounds have clear non-linear momentum dependencies.

Table 14.4: Results of running the analysis method on sub-datasets made by splitting the data into three bins of D^0 momentum. The bias is the difference between the average measured value and the default fit result of 410.0 fs.

	Momentum ranges [GeV/c]		
$\tau(D^0)$ [fs]	$0-37$	$37-53$	$53-1000$
magnet up	412.5 ± 0.8	410.2 ± 0.7	409.5 ± 0.6
magnet down	412.7 ± 0.8	408.5 ± 0.7	408.8 ± 0.6
average	412.6 ± 0.6	409.3 ± 0.5	409.2 ± 0.4
bias	2.6	-0.7	-0.8

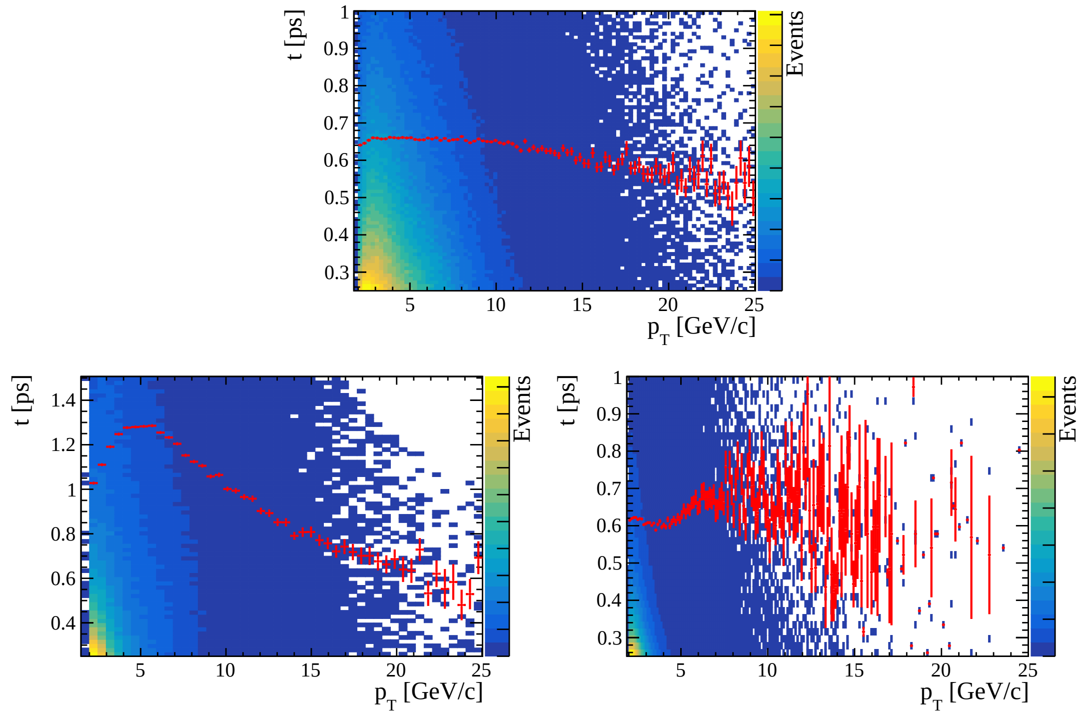


Figure 14.2: Data distributions of transverse momentum against decay-time overlaid with (red points) the profile of decay-time in bins of transverse momentum. All three plots contain data with various cuts applied to $m(D^0)$, Δm and $\ln(\chi^2_{\rm IP})$, to select enriched samples of (top) prompt signal, (bottom left) secondary and (bottom right) remaining backgrounds. All thee components have clear non-linear dependencies against transverse momentum.

Table 14.5: Results of running the analysis method on sub-datasets made by splitting the data into three bins of D^0 transverse momentum. The bias is the difference between the average measured value and the default fit result of 410.0 fs.

	p_T ranges [GeV/c]		
$\tau(D^0)$ [fs]	$0-4.5$	$4.5-7$	$7-9999$
magnet up	413.2 ± 0.7	409.5 ± 0.8	407.6 ± 0.9
magnet down	412.8 ± 0.7	410.3 ± 0.8	–
average	413.0 ± 0.5	409.9 ± 0.5	407.6 ± 0.9
bias	3.0	-0.1	-2.4

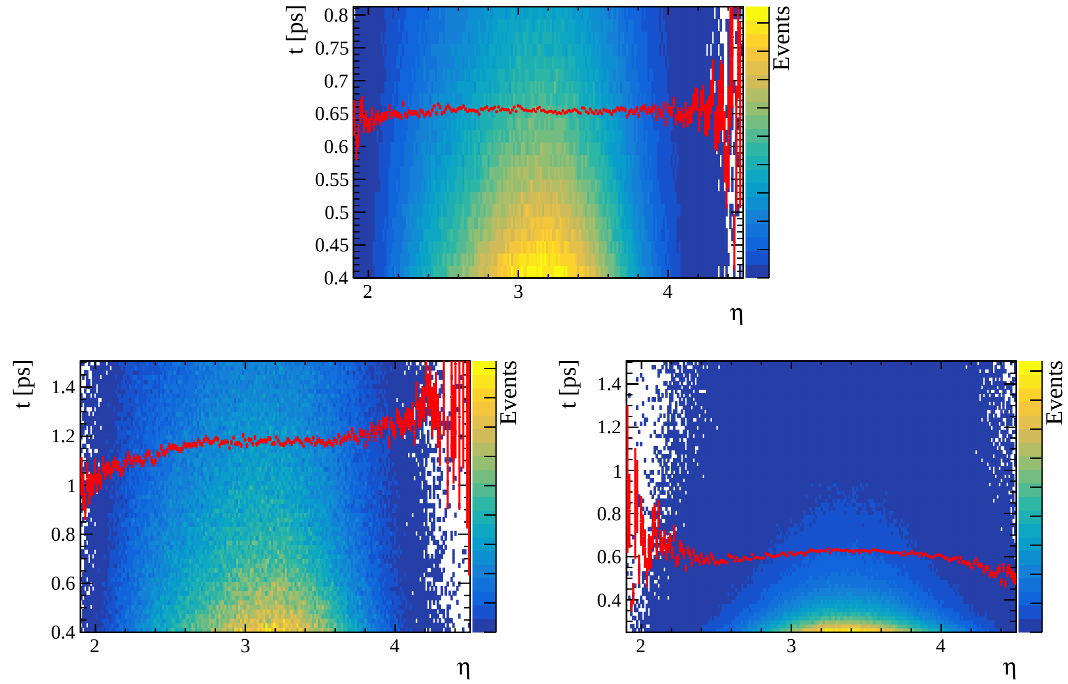


Figure 14.3: Data distributions of pseudorapidity against decay-time overlaid with (red points) the profile of decay-time in bins of pseudorapidity. All three plots contain data with various cuts applied to $m(D^0)$, Δm and $\ln(\chi^2_{\rm IP})$, to select enriched samples of (top) prompt signal, (bottom left) secondary and (bottom right) remaining backgrounds.

The origin of these biases to the prompt decay-time is largely unknown. They may be a result of background contamination, where the decay-time of the backgrounds themselves have kinematic dependencies. Or it may also be due to decay-time acceptance effects created by the selection criteria. The most likely explanation is that the tight selection requirement of $\ln(\chi^2_{\rm IP}) < 1$, used to separate prompt and secondary decays in the tests of this section, introduce a decay-time bias. These tests should be carried out with samples of simulated data to which decay-time biasing selection criteria have not been applied. However, such samples do not yet exist for Run 2. In the absence of a clear explanation, the largest of these biases from the first transverse momentum bin is assigned as an additional systematic uncertainty. It is likely that this assigned systematic is double counting the systematic effects caused by the backgrounds, but it is not clear to what extent this is the case. A conservative approach is favoured when assigning systematics to this measurement feasibility study.

The same tests were done with D^0 η splits of $[1.5, 3, 3.4, 5]$. In that case no significant bias was observed.

14.1.3 Remaining backgrounds

The remaining backgrounds are those from combinatorial and random soft pion events. The systematic uncertainty due to modelling the remaining backgrounds is assigned from the bias in a toy test generated without the secondary background and where the fraction of these remaining backgrounds is varied. The fraction of this background after applying the $\ln(\chi^2_{\mathrm{IP}})$ cut, varies in the data from about 32% in the lowest decay-time bin to 28% in the last bin. The results of this test are shown in Table 14.6. The largest bias of 0.50 fs comes from the test in which 40% of the generated data is from these backgrounds, and is assigned as the systematic uncertainty.

The test that gave the second largest systematic effect was one in which an alternate fit PDF was used for the remaining background component in the Δm fits. Two root exponential PDFs, first introduced in Eq. (7.4), are used instead of one as in the default. The decay parameters of both components are fixed to extreme negative values such that they play no role in the PDFs. A fraction describing the relative size of the two components is allowed to float in the fit. The power parameter of the smaller component fitted to a value of approximately 5.3 in preliminary fits, but was fixed to that value in the final fits to improve stability. The Δm fit quality is not significantly changed by the addition of the new background component, which has a fraction of approximately 0.2 with respect to the first background component in the final fits. The averaged value

Table 14.6: Results of a series of toy tests generated with different remaining background fractions. The first column shows the remaining background fraction, while the second column shows the measured lifetime. The third column shows the lifetime fit $\chi^2_{/\mathrm{ndf}}$ and the last column shows the bias of the measured lifetime with respect to the generated lifetime, which is the PDG value of 410.1 fs.

Background [%]	measured $\tau(D^0)$ [fs]	fit $\chi^2_{/\mathrm{ndf}}$	bias [fs]
25	409.7 ± 0.47	9.5/8	-0.4
35	409.8 ± 0.53	10.7/8	-0.3
40	410.6 ± 0.57	7.4/8	0.5
45	410.3 ± 0.61	7.2/8	0.2
50	410.5 ± 0.67	7.8/8	0.4

from the lifetime fits for this alternate model is (410.4 ± 0.3) fs giving a change in the result of 0.4 fs.

Further tests of the systematic effect of these backgrounds were done using the PDF of Eq. (7.5) for the background component in the Δm fits, using an alternate signal Δm PDF, and by generating further toys while varying the prompt lifetime. These tests gave rise to smaller biases than those already discussed.

14.1.4 Acceptance systematics

As evidenced by the dataset splits, it is not yet clear if the prompt decay-time is unbiased by the selection criteria. To check the magnitude of any decay-time acceptance effects, the analysis method must be repeated with a suitable Run 2 simulation sample, which has the same reconstruction and selection as the 2016 data. With a large simulation sample the decay-time acceptance can be determined and if necessary a correction applied to each bin to account for the effect. This method has the disadvantage that the uncertainty on the final result will be dependent on the statistics of the simulation sample used to determine the correction.

The track finding efficiency of the VELO sub-detector, first discussed in Section 8.3.3, has yet to be determined for 2016 data. Similarly, the upper decay-time acceptance limit arising from the VELO geometry has not yet been investigated. These effects would be accounted for in any correction determined from simulation. However, differences in the number of tracks seen in simulation and real data mean that more reliable results can be obtained using data driven methods and considering each effect separately. These effects create upper acceptance limits at longer decay-times and thus they are unlikely to affect D^0 candidates in the $[0.4, 1.2]$ ps range used in this analysis method.

On the basis of the previous studies in Part II, the largest systematic effects have likely been accounted for in this analysis. However, a full investigation of the outstanding systematic effects is required before this measurement can be unblinded.

14.2 Results

Shown in Fig. 14.4 are the lifetime fits to the magnet polarity up and down sub-datasets, respectively. Poor fit $\chi^2_{/\mathrm{ndf}}$ values would indicate non-linearity in the lifetime fits, which would suggest a problem with the analysis such as

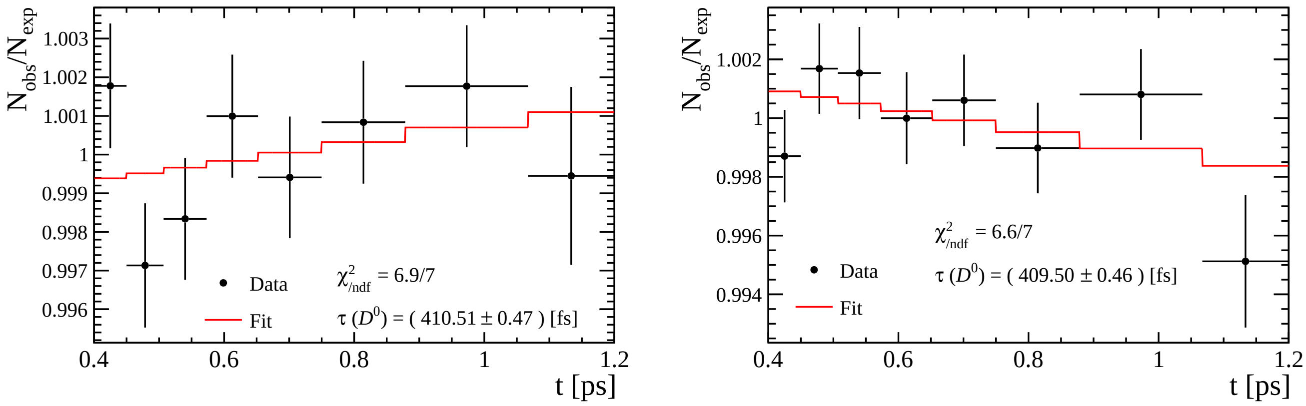


Figure 14.4: Lifetime fits to the (left) magnet polarity up sub-dataset and (right) the polarity down sub-dataset. Shown are (black points) the number of observed events divided by the number of expected events and (red line) the fit function, along with the extracted values of the lifetime parameter and the fit $\chi^2_{/\mathrm{ndf}}$. The horizontal data error bars represent the bin widths.

significant secondary contamination or an acceptance effect. Both fits return good $\chi^2_{/\mathrm{ndf}}$ values and blinded lifetime values that are within the expected range, demonstrating that the analysis is not significantly affected. The average value of the blinded lifetime parameter is given by,

$$\tau(D^0) = (410.0 \pm 0.3 \pm 4.1)\,\mathrm{fs}, \tag{14.1}$$

where the magnet polarity up and down results are weighted using the number of D^0 candidates in each sub-dataset. The magnet polarity samples are compatible at the level of 1.4σ. The sensitivity achieved by this blinded measurement is the second best for this quantity to-date.

The systematic uncertainty on this measurement could be reduced with a clearer understanding of the decay-time dependence on kinematic observables of the data. As discussed, these effects may be caused by background contamination or residual decay-time acceptance effects. Using a suitable simulation sample to check for acceptance effects from the event selection and if necessary correct for them, is the next step for this analysis method.

Chapter 15

Additional methods for secondary separation

Machine learning algorithms (MLAs) may have potential for separating prompt and secondary background events, beyond the $\ln(\chi^2_{\mathrm{IP}})$ requirements used in the previous chapters. A good separation between these classes can be achieved when MLAs are given a number of input variables, each of which has only modest class separation capability. However, these MLA methods can introduce decay-time biases as well as biases to other important variables such as Δm. Biases to Δm should be avoided as this variable is used to separate signal events from the remaining background events. It is necessary to construct an MLA that separates the prompt and secondary classes while also maintaining a flat selection efficiency against t and Δm.

Section 15.1 introduces "decision trees" (DTs), the base classifiers used in the MLAs presented here. Section 15.2 introduces the concept of "boosting", while Section 15.3 describes the "uniform" MLAs that are investigated in this chapter. Due to the lack of Run 2 GEANT4 based simulation samples, the MLAs presented here have not been trained with ideal samples. Instead in Section 15.4, Run 1 simulation is used to train the MLAs and the effect of one of these classifying algorithms on 2016 data is briefly studied.

Presented in Section 15.5 are additional observables that measure the "isolation" of the prompt signal decay in an event. Secondary decays may be distinguishable from prompt decays by the presence of extra tracks coming from the decay of the B meson. The isolation observables studied in this chapter attempt to separate prompt and secondary events by considering observables that depend on extra tracks in the event, near the signal tracks. The use of these observables in uniform MLAs improves the performance of the algorithms.

15.1 Decision trees

A decision tree [70] consists of a series of nodes, each of which has a binary split based on some criterion of the data. All events in a dataset start in the "root" node of the tree. The events are then split into two sub nodes based on the decision of some requirement, $\ln(\chi^2_{\mathrm{IP}}(D^0)) < 2$ for example. The data in each of these sub nodes is then further split based on the requirements of those nodes, which are not necessarily the same. This continues until the data are in one of the "terminal nodes". Each terminal node has an associated signal or background probability and thus the events ending in that terminal node are classified accordingly. In this way, the DT splits a multidimensional space of data features into various rectangular volumes.

The values of the requirements and the features used at each node, often collectively known as the weights, $\boldsymbol{a}$, are determined by "training" the DT with data samples of known class. The weights are chosen to minimise the impurity, given by,

$$\frac{1}{N}\big(n_1 G_1(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{a}) + n_2 G_2(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{a})\big), \tag{15.1}$$

where N is the total number of events in this node, n_j is the number of events at daughter node j, given the current weights, $\boldsymbol{X}$ is the array of data features, $\boldsymbol{y}$ is the vector of known class labels (1 = signal, 0 = background) corresponding to the events in $\boldsymbol{X}$, and G_j is the impurity function at daughter node j. The impurity function is typically given by the "Gini impurity" [70],

$$G_j(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{a}) = \sum_k p_{j,k}(1 - p_{j,k}), \tag{15.2}$$

where $p_{j,k}$ is the fraction of the number of events of class k, in node j. A terminal node is formed when one of the daughter nodes satisfies some criterion, usually when a user defined maximum node depth is reached, or the Gini index is zero. The criterion could also be based on a user defined minimum number of events in the node. The signal probability of each terminal node is determined from the fraction of training events in that node which are signal.

DTs have many flaws, such as their propensity for overtraining and the simplicity of the boundary they construct between signal and background in the multidimensional space of the data features. In most situations, simple rectangular selections are not sufficient to separate classes. For these reasons they are considered "weak classifiers". As discussed in the next section, these classifiers only become powerful tools when multiple DTs are used in an ensemble.

15.2 Boosting and gradient boosting

The technique of "boosting" [70, 90] involves first training a single "base classifier" $h_0(\boldsymbol{x}, \boldsymbol{a}_0)$, where $\boldsymbol{x}$ is a set of values of the input features, the same length as the array of training input features in $\boldsymbol{X}$. The function $h_0(\boldsymbol{x}, \boldsymbol{a}_0)$ is continuous[1] for all $\boldsymbol{x}$ and its weights depend on the training data $\boldsymbol{X}$ and $\boldsymbol{y}$. A series of classifiers $(h_1, \ldots, h_M)$, is then constructed where the per-event weights of incorrectly classified training events are increased for each new classifier. In this way the incorrectly identified events from the preceding classifiers are more likely to be correctly identified by the next classifiers. The final "boosted" classifier is a linear combination of the base classifiers,

$$H_M(\boldsymbol{x}, \alpha_0, \ldots, \alpha_M, \boldsymbol{a}_0, \ldots, \boldsymbol{a}_M) = \sum_{m=0}^{M} \alpha_m h(\boldsymbol{x}, \boldsymbol{a}_m), \tag{15.3}$$

where $\{\alpha_0, \ldots, \alpha_M\}$ are the linear coefficients of the ensemble that are determined by some loss function, $L(\boldsymbol{y}, H_M)$. One example of this type of MLA is the "adaptive boost" (AdaBoost) algorithm [91]. The per-event weights of the misidentified events are multiplied by $e^{(\alpha_{m-1})}$ before training base classifier m, and an exponential loss function gives the ensemble coefficients as,

$$\alpha_m = \ln\left(\frac{1 - \delta_{m-1}}{\delta_{m-1}}\right), \tag{15.4}$$

where δ_{m-1} is the fraction of misidentified events in the previous classifier.

The idea behind "gradient" boosting [92, 93] is to treat the base classifiers as vectors in some space, determine the gradient of the loss function in this space and move along this direction to find the minimum of the loss function. In this way gradient boosting is analogous to the steepest descent numerical minimisation method and thus allows any differentiable loss function to be used for boosting. The set of gradients of the loss function in this space is given by,

$$\boldsymbol{g}_m(\boldsymbol{x}_i) := \left\{\frac{\partial L(y_i, H_M(\boldsymbol{x}_i, \ldots))}{\partial H_M(\boldsymbol{x}_i, \ldots)}\right\}_{i=1}^{N} := \{g_m(\boldsymbol{x}_i)\}_{i=1}^{N}, \tag{15.5}$$

where the $\boldsymbol{x}_i$ denote the individual sets of input features for each of the N training data events in $\boldsymbol{X}$. These gradients are only determined for each of the i training events, so a new classifier is constructed that best fits the direction defined by the negative gradient, $-\boldsymbol{g}_m$.

[1] Or piece-wise continuous

What does fitting the new DT to a negative gradient in function space require in practice? The predictions of the new DT for each training event must match the components of the negative gradient so that when they are added to the ensemble, the total prediction of all events move towards the true predicted value. As an example consider an event with input feature values, $\boldsymbol{x}_j$, with true class label $y_j = 1$ (signal) that has a prediction of $h_0(\boldsymbol{x}_j) = 0.75$ from the zeroth DT. If a least squares loss function is used, the next DT in the ensemble aims to add $(1 - 0.75)$ to the prediction of the ensemble. Thus the new prediction that the DT must target for this event is $h_1(\boldsymbol{x}_j) = 0.25$. In this case 0.25 is the negative gradient of the loss function evaluated for training event $\boldsymbol{x}_j$. If h_1 under performs for this event and only returns 0.1 say, then the next DT must target a prediction of $h_2(\boldsymbol{x}_j) = 0.25 - 0.1$. In this way, the direction of the new DT in function space matches closely to the direction of the negative gradient. So in practice, the new DT is fitted to the direction of the negative gradient by using the negative gradients as the true prediction (y_j) for each event, which are input to the new DT when training.

The true labels $\boldsymbol{y}$ are included in the loss function, thus the negative gradients encode the y_i information and are called “pseudo-responses” in the literature. The new DT, with weights determined using the pseudo-responses, is the next element in the ensemble, $h_m(\boldsymbol{x}, \boldsymbol{a}_m)$.

Note that in the case of gradient boosting there are no ensemble coefficients α_m, so the total classifier is just the sum of the individual classifiers. A “learning rate” parameter can be given to the algorithm that scales the prediction from each DT by a fixed amount. This effectively scales the step size in the direction of the negative gradient. It determines how fast the classifier moves towards the minimum, and how much influence a single classifier can have on the total prediction.

15.3 Uniform boosting

An MLA that separates component classes while keeping a flat efficiency against user defined “special features” is called a “uniform” MLA. The special features used in the analyses to follow are the D^0 decay-time, t and the mass difference Δm. The MLAs presented here are based on either boosted decision trees (BDTs) or gradient BDTs (GBDTs), introduced in previous sections.

The loss functions used with gradient boosting for this analysis are modified so that the resulting MLAs have uniform efficiencies against the chosen special features. The technique for building uniform GBDTs is to use a linear combination of two loss functions,

$$L = L_{\text{separation}} + cL_{\text{uniformity}}. \tag{15.6}$$

In this way, the loss function separates classes but also penalises the total function for non-uniform responses, where the "uniformity coefficient" c, controls the trade-off between separation and uniformity.

The two uniform GBDTs used in this analysis are uGB+knnFL and uGB+binFL [94], for which the AdaLoss function,

$$L_{\text{Ada}}(\boldsymbol{X}, \boldsymbol{y}, r) = \sum_{i=1}^{N} \exp\left(-y_i r(\boldsymbol{x}_i)\right), \tag{15.7}$$

is used for separation. Here $\boldsymbol{X}$ is the array of training data events and input features. The symbol $\boldsymbol{x}$ denotes a set of values of input features and with an index subscript i it denotes one of the N training data events, for which the class label y_i is known. The function $r := r(\boldsymbol{x}) = H_{m-1}(\boldsymbol{x})$ (see Eq. (15.3)) is the response of the total MLA ensemble from the previous boosting step, defined for all possible sets of input values $\boldsymbol{x}$. The uniform component of the loss function for uGB+binFL is constructed by first binning the special uniform features into b bins. If the distribution of the classifier response in each bin, $f_b(r)$, is the same as the global response distribution, $f(r)$, then any selection requirement on the response will not bias the uniform features. A loss function is constructed by comparing these distributions, namely their cumulative distributions, $F(r)$,

$$L_{\text{binFL}} = \sum_b w_b \int |F_b(r) - F(r)|^2 dr,$$

where w_b is the fraction of signal events in bin b. In this way, the difference between the binned distributions of the response and the total distribution of the response is minimised when minimising the total loss function, L.

The second MLA of this type, uGB+knnFL uses the same uniformity loss function but the binned classifier response is replaced with the k nearest neighbours (knn) response. The knn events in the space of special features are grouped together and their response is compared to the total response. In this case the bin weight is replaced with the fraction of signal events in the knn sample.

The BDT approach to uniformity is to update the per-event training weights for each new boost to ensure uniformity. Events in regions where the efficiency is greater(less) than the mean are given smaller(larger) weights. The usual boosting weights are multiplied by a uniformity weight, $b'^{i}_{m} = u^{i}_{m} b^{i}_{m}$. The algorithm of this type considered for this analysis is uBoost described in [95], which uses the AdaBoost [91] loss function to determine the classification boosting weight. The uniformity boosting weight, u^{i}_{m}, is determined from minimising a loss function constructed from the difference between the mean efficiency $\bar{\epsilon}_m$ given some response requirement, and the local efficiency ϵ^{i}_{m} for event i calculated from its knn sample. Hence this must be calculated for many different requirements on the classifier response.

15.4 Training

When training an MLA, a sample of signal and background data is supplied. Those samples are split into two groups. The first group is used to train the MLA, while the second statistically independent "test" group is used to evaluate the performance of the algorithm. The figures shown in this section are produced with the testing group, unless stated otherwise.

The prompt and secondary Run 1 simulation samples described in Section 12.3 are used here to train the algorithms as a demonstration because more suitable samples are not available. This simulation contains significant decay-time biases due to the selection criteria used. Thus the correlations between the input features and the special features will be different in the training data and the real 2016 data. As a result, an MLA that produces a flat efficiency in the special features of this simulated data is unlikely to produce a flat efficiency in the real 2016 data. Consequently, the results shown here can only be taken as a proof of principle. The decay-time biases are smaller in the Run 2 dataset than in Run 1 simulation, and so the aim of keeping the efficiency of the special features flat with respect to the MLA response is more difficult in the Run 1 simulation used here.

Two uniform MLAs were considered in these studies, uGB+knnFL and uGB+binFL. The uBoost algorithm was also considered, but found to have reduced performance for this use case. None of the MLAs investigated successfully produced flat efficiencies in the special features, t and Δm. However, flat efficiencies were achieved by using additional parameters, as described in Section 15.5. Of all the MLAs investigated here the best performing one was the uGB+knnFL algorithm with the uniformity coefficient, $c = 2$ and 300 trees in the boosted ensemble.

All algorithms have a max depth of 4, which is the maximum number of nodes between the root and any terminal node of a DT. It was found that lower values

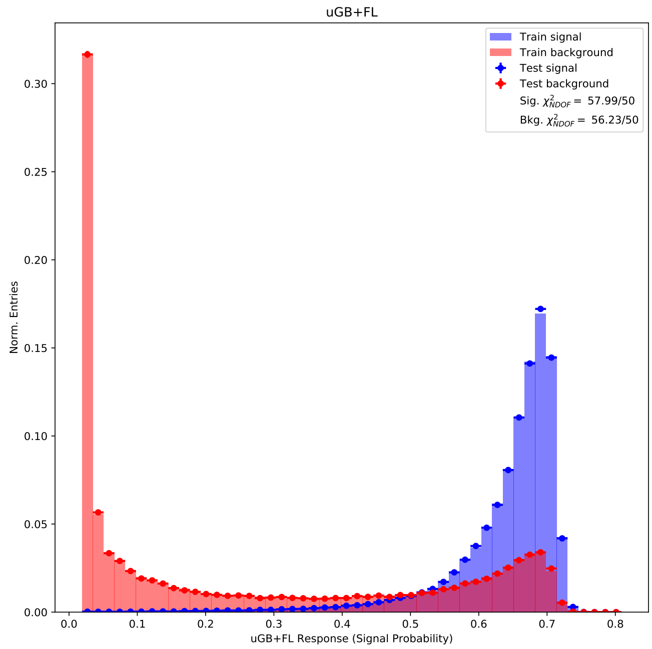


Figure 15.1: Overtraining χ^2 test for the best MLA showing the MLA response for the signal and background, training and testing Run 1 simulation samples. Shown are the (filled) training data group and (points) the testing group for (red) the background sample and (blue) the signal sample. The χ^2 test indicates that the distributions are consistent and thus there is no overtraining.

of this parameter result in larger biases of the special features while higher values result in overtraining. Overtraining results when the MLA learns to separate the signal and background samples by their statistical fluctuations, and thus has a good performance on those samples, but performs inefficiently on real data. This usually happens when the MLA has too many constituent trees. To test for overtraining a χ^2 test comparing the MLA response from the training sample and the testing sample is performed such that a poor χ^2 value indicates overtraining. Fig. 15.1 shows the overtraining χ^2 test for the best MLA indicating that the MLA is not overtrained.

The number of trees used in these studies is 100 or 300. Trials with less than 100 trees were found to be sub-optimal, while MLAs with more than 300 trees gave no improvement in performance, but significantly increased the amount of time needed for training. The number of trees and the performance of the MLA are closely related to the learning rate (see Section 15.2). Decreasing the learning rate while keeping the number of trees fixed has the same effect on performance as reducing the number of trees. A value of 0.1 was chosen for the learning rate. Using 0.5 as a trial made little difference to the performance or the uniformity of the resulting MLAs.

The final parameter to be set is the uniformity coefficient, c. As can be expected, small values of c (below 0.05) result in MLAs with performance comparable to non-uniform MLAs. Values been 0.7 and 5 were investigated for these studies. Low values near 0.7 had reduced performance from non-uniform MLAs and didn't significantly change the uniformity of the special features. High values above 3 were very successful at removing biases in one of the special features, namely decay-time, but introduced biases into the other feature, Δm. The Δm biases in the training data are larger than the decay-time biases, which may be the cause of this effect. It was found that c values of about 2 or 2.5 are a compromise between good performance and uniformity, without introducing Δm biases.

Shown in Table 15.1 are the feature importances of the input observables, which indicate the importance of a feature for separating the classes and keeping the special feature efficiencies flat. This metric is calculated from the Gini index of every node where a certain feature is used. It is thus a measure of the separation of the two training classes given the selection requirements used at each node in the trained MLA. As the uniform MLAs disfavour features that bias the special features, those that create biases have their importances reduced. The dist(D^*) observable is the z distance between the PV and the D^* decay vertex. For prompt events these should be the same vertex and dist(D^*) should be a resolution function. For secondary events the two vertices are not the same due to the intermediate B meson track, and the observable is a resolution function plus some long positive tail. The Δm_{DTF} observable is constructed with the decay tree fitter (DTF) algorithm [96], which applies constraints to the variables when fitting the tracks and the vertices in a chain of decays. The D^0 mass is constrained to the PDG value and the daughter tracks are constrained to exactly coincide with the D^0 decay vertex. Thus the D^0 decay vertex has no uncertainty and the resolution of observables such as Δm is reduced. The observable Δm_{DTF} is the difference between the D^* mass with and without these constraints.

The efficiency of the special features with various selection requirements applied to the best MLA response are shown in Fig. 15.2. Choosing a value of the MLA response, such that events below the value are not considered, is equivalent to choosing a signal efficiency. Similarly, a chosen signal efficiency can be associated

Table 15.1: Feature importances of the best performing MLA, which used the uGB+knnFL algorithm with a flatness coefficient of 2 and 300 trees. Here FD is the flight distance.

Observable	Importance
$\ln(\chi^2_{\mathrm{IP}}(D^*))$	0.30
$\ln(\mathrm{IP}(D^0))$	0.18
$\ln(\chi^2_{\mathrm{FD}}(D^*))$	0.13
$\ln(\chi^2_{\mathrm{FD}}(D^0))$	0.10
Δm_{DTF}	0.06
$\ln(\chi^2_{\mathrm{IP}}(D^0))$	0.05
$\mathrm{dist}(D^*)$	0.05
$\ln(\mathrm{IP}(\pi_s))$	0.04
$\ln(\mathrm{FD}(D^0))$	0.04
$\ln(\chi^2_{\mathrm{IP}}(\pi_s))$	0.02
$\ln(\mathrm{FD}(D^*))$	0.02
DIRA D^*	0.01

with a selection requirement on the MLA response as shown in the figure legends. Events with an MLA response value greater than 0.7070 are chosen, which gives a signal efficiency of about 10%. As mentioned, this MLA is not able to flatten the efficiency of the special features.

The effectiveness of an MLA at separating the signal and background classes is determined from a receiver operating characteristic (ROC) curve, which in this case is simply the signal efficiency against the background rejection efficiency. In this analysis the percentage of secondary contamination must be as low as possible to reduce systematic biases to the lifetime. For this reason, extreme selection requirements must be applied to the MLA response. Fig. 15.3(left) shows a zoom of the ROC curve for the best MLA, in the region of interest where most of the secondary background can be removed. Choosing 10% signal efficiency gives a secondary background rejection of 98.5%, which is a $10:1$ reduction in the signal and a $66:1$ reduction in the secondary contamination. The scale of this reduction is considered to be sufficient for reducing systematic biases from secondary decay contamination.

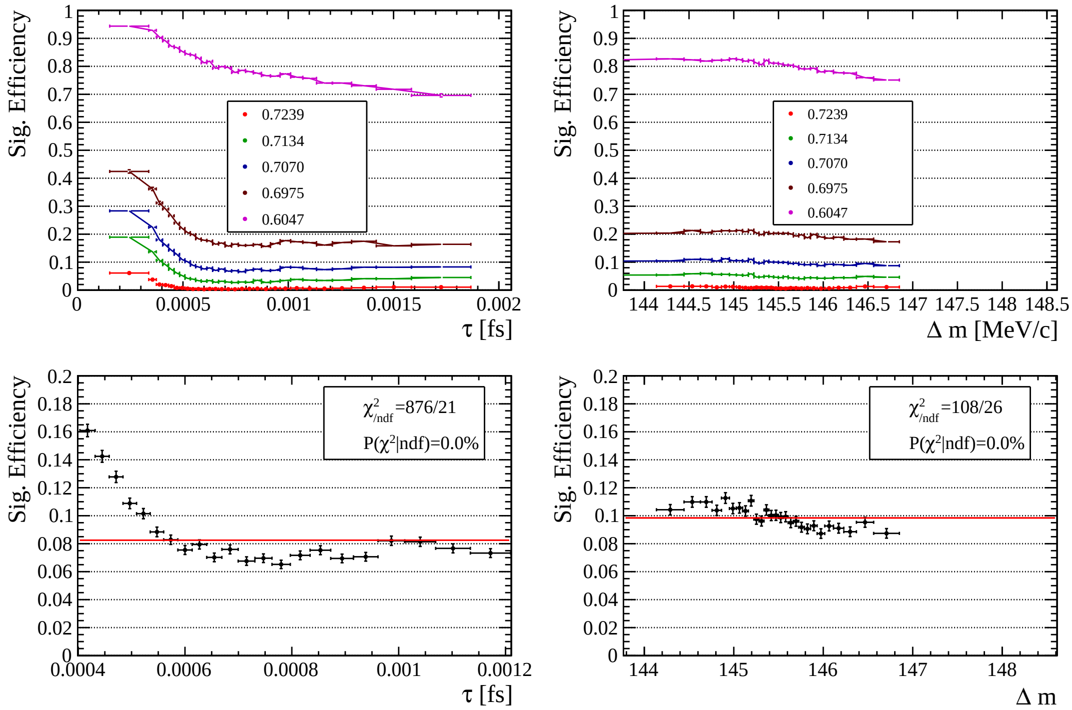


Figure 15.2: Uniform feature efficiencies with various selection requirements applied to the MLA response from the best performing MLA. Shown are (left) the decay-time efficiencies and (right) the Δm efficiencies. The top plots show the efficiencies for a range of MLA response requirement values, which correspond to roughly 80%, 20%, 10%, 5% and 1% signal efficiencies. The corresponding MLA response requirement values are shown in the legends. The bottom plots show fits to the 10% efficiency distributions to check for uniformity, showing that this MLA does not succeed in flattening the efficiency.

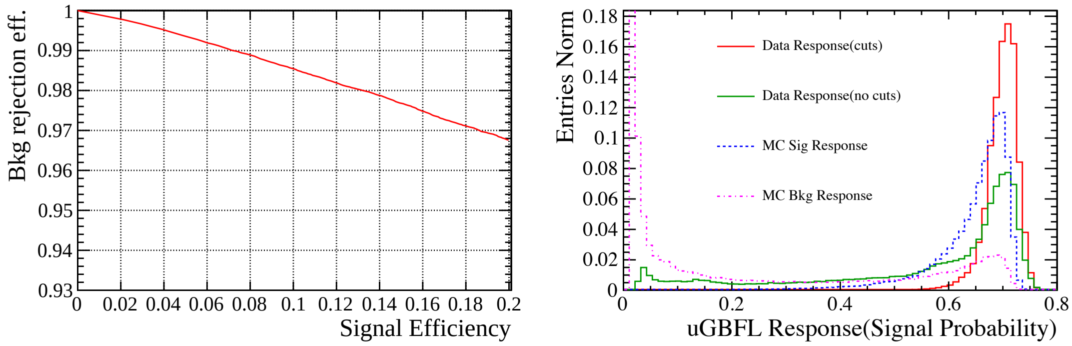


Figure 15.3: Shown is (left) the ROC curve for the testing simulation group of the signal efficiency against the background rejection efficiency for the best MLA and (right) a comparison of the MLA response for real 2016 data and Run 1 simulation.

The MLA responses of the best MLA and the real 2016 data are shown in Fig. 15.3(right). Both the signal and background responses of the testing data are shown along with the data, with and without requirements to select a relatively pure sample of prompt signal decays (selection requirements were applied on $m(D^0)$, Δm and $\ln(\chi^2_{\text{IP}})$). As is to be expected, the samples used for training and the real data have very different responses when classified by the MLA. The residual secondary contamination after applying an MLA response requirement can be estimated from the ROC curve. However, the peak positions of the MLA response for the data (red line) and prompt signal simulation (blue dashed line) are different. Therefore, if an MLA response requirement is applied to the data, the residual secondary contamination in the data will be different from that estimated from the ROC curve. This is due to the significantly different acceptance effects and feature correlations between the real 2016 data and the Run 1 simulations, and by consequence, makes the MLAs shown here unsuitable for 2016 data analyses.

15.5 Decay chain isolation

This section introduces observables that measure the "isolation" of the prompt signal decay from other tracks in the event, and uses them to improve the uniform MLAs of the previous section.

As mentioned in Section 3.2.4, the turbo trigger lines only contain information about the signal tracks in an event. Thus, the information needed to construct isolation observables is not present in the Run 2 data considered in the previous chapters. Instead, samples of $D^0 \to K^-\pi^+$ Run 1 simulation are used, which are similar to those used in the A_Γ analysis of Part II. A Run 1 data sample with the same selection as in Part II is also used to illustrate the distribution of the isolation observables in real data. In the data sample the D^0 comes from the tagging decay $D^{*+} \to D^0\pi^+$, which can give some separation between prompt and secondary events. Hard PID requirements of PIDK > 5 for the daughter kaon, and PIDK < -5 for the pion, are applied. The number of D^0 candidates in each event is restricted to be one, and the lifetime must lie in the range $[0.25, 10]$ ps. In addition to these requirements, the $m(D^0)$ and Δm of the data must lie in the ranges $[1860, 1890]\,\text{MeV}/c^2$ and $[143, 147]\,\text{MeV}/c^2$, respectively, to reduce combinatorial and random soft pion backgrounds. The total number of events in each sample is 811,000 for the prompt simulation, 173,000 for the secondary simulation and 927,000 for the data.

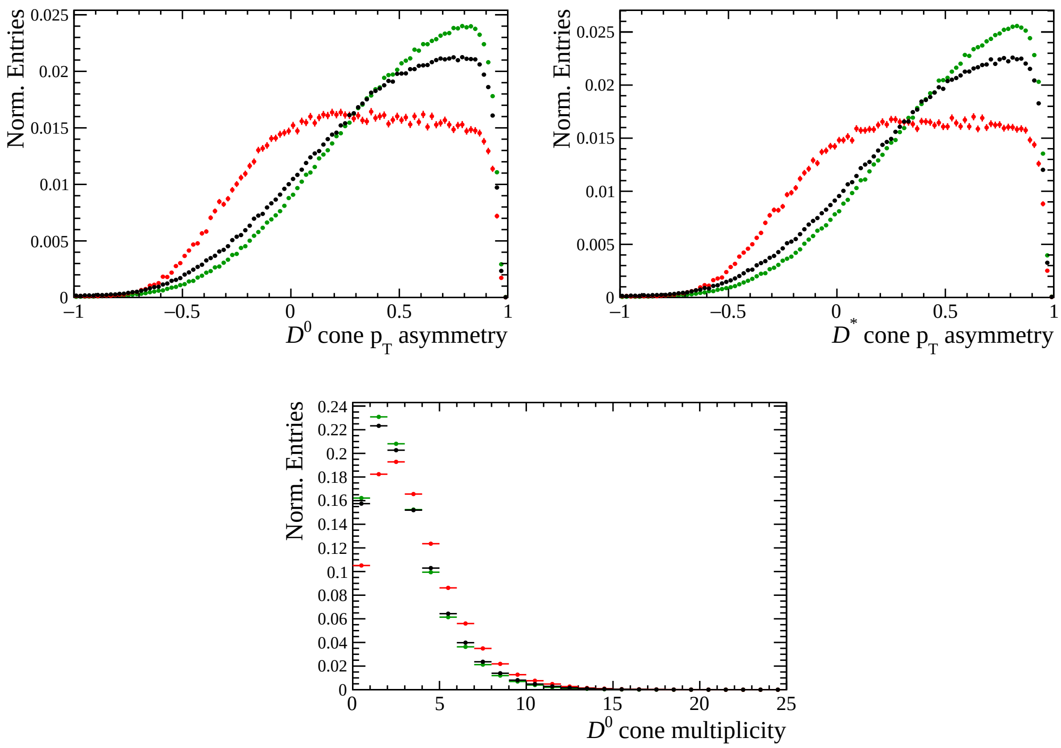


Figure 15.4: Observables constructed from the isolation cones described in the text. Shown are the normalised distributions of (top left) the p_T asymmetry between the D^0 and the isolation cone constructed around it, (top right) the p_T asymmetry between the D^* and the isolation cone constructed around it and (bottom) the multiplicity of the D^0 isolation cone. The colours represent (green) the prompt simulation sample, (red) the secondary simulation sample and (black) the data sample.

Isolation observables

Isolation observables constructed from a cone surrounding the signal track are the first group of observables considered [2]. The momentum direction of the mother particle of the decay is compared to the momentum direction of every other particle in the event, excluding the other tracks that form the signal decay. In the case studied here the mother particle is either the D^* or D^0. The η and ϕ angle differences, $\Delta\eta$ and $\Delta\phi$, between the mother particle momentum and a candidate track are calculated. If the candidate satisfies the relation,

$$\sqrt{\Delta\phi^2 + \Delta\eta^2} < \Delta r, \tag{15.8}$$

then its momentum is added to the cone surrounding the mother particle. The Δr criterion is defined by the user. It was found that $\Delta r = 0.5$ gave the best

[2] These observables are calculated using the DaVinci tool `TupleToolConeIsolation`

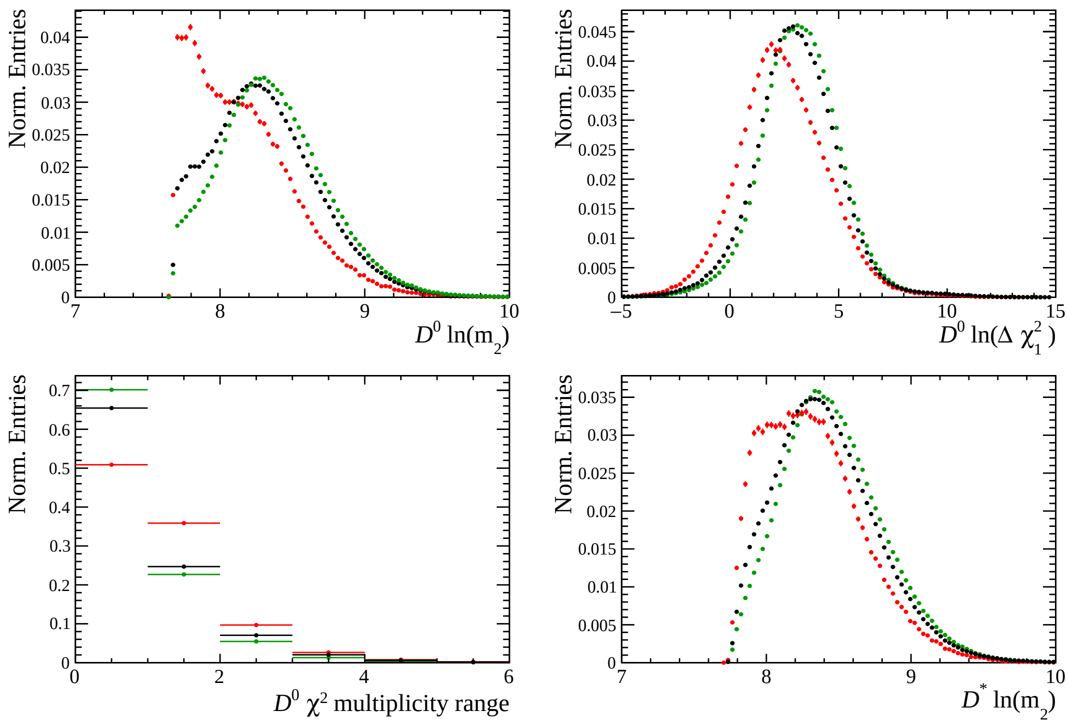


Figure 15.5: Observables constructed from the adding of extra tracks to vertices as described in the text. Shown are the normalised distributions of (top left) the natural logarithm of the two extra track mass for the D^0 vertex, (top right) the natural logarithm of the one track minimum χ^2 difference, (bottom left) the multiplicity range as described in the text, and (bottom right) the two extra track mass for the D^* vertex. The colours represent (green) the prompt simulation sample, (red) the secondary simulation sample and (black) the data sample.

separation between prompt and secondary in the observables considered. Cone sizes of 0.25 to 3 in step sizes of 0.25 were also considered. Only cones consisting of charged tracks were considered, as cones filled with neutral particle tracks did not show significant separation between simulated prompt and secondary events.

Three of these cone isolation variables have significant differences in simulated prompt and secondary events. These are the p_t asymmetry between the mother particle and the resulting isolation cone for both the D^0 and the D^*, and the multiplicity of the D^0 isolation cone. Fig. 15.4 shows the distributions of the simulation samples and the data in these variables.

The second group of observables considered are vertex isolation observables[3]. These are constructed by adding extra tracks to the vertex fits and comparing the fit χ^2 before and after. The "one track" observable, $\Delta\chi^2_1$, is the χ^2 difference

[3]These observables are calculated using the DaVinci tool `TupleToolVtxIsoln`

Table 15.2: Feature importances of control MLAs without isolation observables for three trainings with different flatness coefficients. Here FD is the flight distance.

Observable	$c = 1.9$	$c = 2.5$	$c = 5$
$\ln(\chi^2_{\mathrm{IP}}(D^0))$	0.39	0.24	0.13
$\ln(\mathrm{IP}(D^0))$	0.36	0.35	0.19
$\mathrm{dist}(D^*)$	0.08	0.18	0.52
$\ln(\chi^2_{\mathrm{FD}}(D^*))$	0.08	0.09	0.02
$\ln(\chi^2_{\mathrm{IP}}(D^*))$	0.04	0.04	0.03
$\ln(\chi^2_{\mathrm{IP}}(\pi_s))$	0.04	0.03	0.01
$\ln(\mathrm{FD}(D^*))$	0.02	0.06	0.11

when adding the single track that gives the smallest change in the fit χ^2. Similarly, the "two tracks" observable, $\Delta\chi^2_2$, is the difference in vertex fit χ^2 when adding two tracks to the vertex fit, that give the smallest change in χ^2. The mass of the resulting composite particle decaying at the vertex, $m_{1(2)}$, is also calculated when adding one(two) extra track(s) to the vertex.

There are four of these observables that show significant differences between the prompt and secondary simulation samples. For D^0 the number of tracks that can be added to the vertex while the χ^2 remains within a given range is significantly smaller for prompt events, referred to as the χ^2 "multiplicity range" in what follows. The single largest difference is seen in the mass of the composite particle when adding two tracks for the D^0 vertex. Similarly, the D^* version of this observable also shows a significant prompt and secondary difference. The final observable considered in this chapter is the one track $\Delta\chi^2_1$ for the D^0. Fig. 15.5 shows the distributions for these four observables for prompt and secondary simulation samples and for the data sample.

MLA study

The MLA used for this study is uGB+knnFL as introduced in Section 15.3. The prompt and secondary simulation samples are used to train the MLA, both with the isolation observables and without, as a control. A number of different uniformity coefficients are trialled in each case. The input features used for the test and control trainings are shown in Table 15.2 along with the feature

Table 15.3: Feature importances of the MLA with isolation observables that has the best performing training. Here FD is the flight distance.

Observable	$c = 2.5$
$\ln(\chi^2_{\mathrm{IP}}(D^0))$	0.27
$\ln(\mathrm{IP}(D^0))$	0.27
$\ln(\chi^2_{\mathrm{FD}}(D^*))$	0.10
$\mathrm{dist}(D^*)$	0.07
D^0 p_T asym.	0.05
D^0 cone muli.	0.04
$\ln(\chi^2_{\mathrm{IP}}(D^*))$	0.04
D^* p_T asym.	0.03
$\ln(\mathrm{FD}(D^*))$	0.03
D^* $\ln(m_2)$	0.03
D^0 $\ln(\Delta\chi^2_1)$	0.02
$\ln(\chi^2_{\mathrm{IP}}(\pi_s))$	0.02
D^0 $\ln(m_2)$	0.02
D^0 χ^2 mult. range	0.01

importances of three control trainings without isolation variables. These three are chosen as the first two have the minimum (1.9) and maximum (5) values of the uniformity coefficient, and the third has the same uniformity coefficient (2.5) as the best training that uses the isolation observables. The DTF variables used in the MLAs of previous sections are not present in these simulation samples and so could not be used as training features. The isolation variables used are shown in Table 15.3 along with the feature importances for the best preforming MLA with a uniformity coefficient of $c = 2.5$. The isolation variables appear to only play a small role in the separation of the datasets, but they are very effective at improving the uniformity of the special variables, as will be shown. This may in part be due to the reduced feature importance of the biasing IP observable.

All the MLAs trained without the isolation observables produced biased decay time and/or Δm distributions in the testing sample. Almost all of the test MLAs, trained with the isolation observables, had biased decay times for all selection requirements on the resulting MLA response. However, the decay times are not biased across the entire range, but only in the region below 0.4 ps, and in

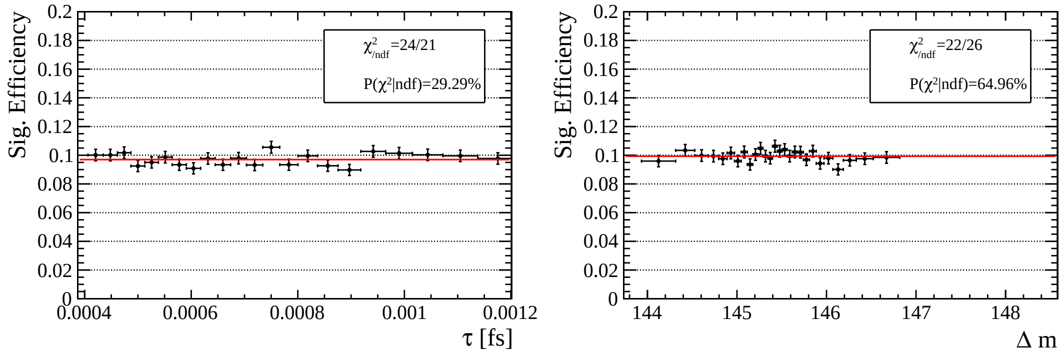


Figure 15.6: Signal efficiencies for the best test MLA with isolation variables. The MLA response requirement that produces this result is > 0.726. The distributions shown are the efficiencies against (left) the decay time in the range $[0.4, 1.2]$ ps and (right) the Δm.

some cases the region above about 1.2 ps. Thus, choosing the decay-time region $[0.4, 1.2]$ ps results in test MLAs that do not bias the special features. It is not clear that the addition of the isolation variables fixes the decay-time biases. It may be the case that the addition of any extra variables with modest separation of prompt and secondary will help to reduce the biases, by reducing the MLAs reliance on IP. In the reduced decay-time range, where acceptance effects are expected to be smaller, the control MLAs without isolation variables are still decay-time biasing.

MLA response selection requirements resulting in 10% and 20% prompt signal efficiency are considered in the reduced decay-time range. Only one MLA, which had $c = 3$, produces uniform special features with an MLA response requirement resulting in a 20% signal efficiency. Three of the MLAs produce uniform special features with MLA response requirements that result in a 10% signal efficiency, these were $c = 2, 2.1$ and 2.5; all have a similar secondary rejection. The prompt signal efficiencies against decay time and Δm for the MLA with $c = 2.5$ are shown in Fig. 15.6, along with a linear fit and its χ^2. By applying this MLA response requirement, 10% of the prompt signal is retained and about 98.8% of secondary events are removed. The scale of this reduction is considered to be sufficient for reducing systematic biases from secondary decay contamination.

The use of one of these isolation variables in a cut based analysis is investigated in Appendix D. Selection requirements are applied to $\ln(\chi^2_{\mathrm{IP}})$ and the two track mass for the D^0, m_2, to remove a large fraction of secondary background.

15.6 Conclusion

The use of uniform MLAs for the separation of prompt and secondary events, without introducing biases to the special t and Δm variables was presented in this chapter. Due to the lack of suitable training data, the studies shown are only a proof of the principle of this method. Initial attempts at constructing trained algorithms that keep the t and Δm efficiencies flat against the MLA response were unsuccessful. However with the addition of further separating features, namely the isolation observables, flat efficiencies were obtained.

As mentioned, the isolation variables cannot be determined in the current Run 2 dataset, namely the turbo stream data. A new stream similar to the turbo stream is planned for future data taking at LHCb. This stream is called turbo selective persistence (TurboSP), and will allow some extra information to be stored in the output data files. The tracks required for the isolation cones and for producing minimum changes in vertex fits' χ^2 can similarly be added to the output data files.

The uGB+knnFL algorithm used in this chapter shows potential for improving the separation of prompt and secondary events in future data that will be taken at LHCb. The additional use of isolation observables also shows potential, but the minimal gain in separation and the time commitment needed to add these observables to the TurboSP stream mean that it may not be worth while. However, in the absence of other separating features, these observables will be necessary for effectively separating prompt and secondary while ensuring a uniform efficiency of decay-time and Δm.

Chapter 16

Prospects of new methods

The blinded measurement of the D^0 lifetime presented in this part has the second best sensitivity of any single measurement of this quantity to-date. A number of additional checks must be carried out before the measured value can be unblinded, and a suitable sample of Run 2 simulation is needed for cross-checks and for determining an acceptance correction if one is required. Consequently the study is presented here as a feasibility study rather than a measurement.

An issue to be resolved with this measurement method is the systematic bias caused by secondary contamination. A $\ln(\chi^2_{\text{IP}})$ requirement does not sufficiently reduce secondary contamination for a direct measurement such as this. The secondary background is reduced to approximately 3% by this requirement, which still returns a large systematic bias. A potential solution to this issue was discussed that uses MLAs to separate prompt signal and secondary background events. The percentage of secondary events that can be removed using these algorithms has been shown; however, the uniformity of the special decay-time and Δm features is not guaranteed. These MLA algorithms may be able to separate prompt and secondary events if an appropriate training sample can be produced, in which the correlations between features are the same as in real 2016 data.

The addition of extra observables that measure the isolation of the signal decay within a collision event was also investigated. These isolation observables do not significantly improve the separation of prompt and secondary events, but they reduce the MLAs reliance on the decay-time biasing IP features, such that the special features are not biased by the MLA. However, the current LHCb Run 2 Turbo stream data does not store the necessary information to determine the isolation variables. The introduction of the TurboSP stream is necessary to persist the information for calculating the isolation observables.

Methods such as those described in this part, which sacrifice systematic accuracy for speed and simplicity, may be necessary in future analyses at LHCb. Larger datasets, increasing rates of data taking and a relative decrease in storage and computing power will necessitate long term planning of analyses. The stored data will have to be smaller and only contain the essential information for an analysis, and the methods must be fast enough to make use of the incredible amount of statistics available while being clever enough to reduce systematic effects.

Part IV

End matter

Chapter 17

Final conclusion

The overall purpose of studies presented in this thesis is to improve the understanding of *CP* violation, due to its potential importance to the matter-antimatter asymmetry in the universe. This is done through the measurement of the *CP* observable A_Γ in the charm sector using LHCb data. The final results presented in this thesis are,

$$\begin{aligned} A_\Gamma(K^+K^-) &= (-0.14 \pm 0.37 \pm 0.10) \times 10^{-3}, \\ A_\Gamma(\pi^+\pi^-) &= (\ \ 0.14 \pm 0.63 \pm 0.15) \times 10^{-3}, \end{aligned}$$

which can be averaged to give,

$$A_\Gamma = (-0.07 \pm 0.32 \pm 0.11) \times 10^{-3}.$$

This is of similar precision to the other measurement of A_Γ presented in the same LHCb paper [66],

$$A_\Gamma = (-0.13 \pm 0.28 \pm 0.10) \times 10^{-3},$$

which was obtained using a complementary technique. This is the world's best determination of A_Γ and is compared to other results in Fig. 10.8(right) in Chapter 10.

The value of A_Γ is consistent with zero and thus with *CP* symmetry in the charm sector. No *CP* violation has been observed in this sector by this or any other experimental measurement. This measurement will continue to have a significant effect on the average calculated values of important charm mixing and *CP* parameters for at least ten years. The effect of these measurements on the parameters ϕ and $|q/p|$ has been shown in Chapter 10.

A number of outstanding problems with the method to measure A_Γ were discussed in Part III. These problems include, the CPU requirements of the swimming algorithm, the storage requirements for large data files, the time requirements for running unbinned maximum likelihood fits, and residual decay-time acceptance effects. The methods presented in this thesis, using the blinded measurement of the D^0 lifetime as an example, attempt to ameliorate these problems while reducing decay-time biases from event selections. The blinded result presented in this thesis, from the method presented in Chapter 13, has a precision of,

$$\tau(D^0) = (\mathrm{xxx} \pm 0.3 \pm 4.1)\,\mathrm{fs}, \tag{17.1}$$

which has the second best precision of any single measurement of this quantity to-date. A clearer understanding of the systematic uncertainties could significantly improve the impact of this measurement. The method presented relies heavily on the new trigger strategy of LHCb in more recent years of data taking, and on the specially designed triggers that reduce decay-time biases. A comparatively simple method has been presented that solves the problems of time and CPU usage, but has not yet been definitively demonstrated to fix the decay-time biases. Suitable simulation samples are needed to check the new method for decay-time biases and to correct for them if necessary. Such simulation samples will be available within a few months.

Future measurements of A_Γ will be made by the Belle II and LHCb experiments. A measurement at the soon-to-start Belle II experiment will have a sensitivity of 0.3×10^{-3} [97], which will be an important statistically independent addition to the world average. At LHCb a measurement is currently underway using the 2015 – 2017 data sample and a binned method. Such a measurement may achieve a sensitivity of about 0.16×10^{-3}. The analysts are using the conventional triggers as opposed to the lifetime unbiased triggers, so there may also be an A_Γ measurement with the unbiased triggers using a method similar to that presented in this thesis. An LHCb measurement with the semileptonic tagged secondary sample from the newest data can also be done.

At LHCb a measurement is underway to determine y_{CP} from the semileptonic tagged Run 1 secondary samples of $D^0 \to K^-\pi^+$. Similarly, the y_{CP} measurement using the prompt Run 1 sample is underway. That analysis uses the swimming algorithm and a similar method to the unbinned maximum likelihood analysis of this thesis. It's not yet clear if the difference in acceptance between the different final states can be resolved in Run 1 data with a sufficiently small systematic uncertainty. Going forward to Run 2, if tests on the simplified method presented in this thesis are promising, then the new method can be used to measure y_{CP}. If the method can remove enough of the decay-time biases such that a direct lifetime measurement can be made, then it will be able to measure y_{CP} without great difficulty.

Looking further to the future of LHCb, A_Γ and y_{CP} will be measured with improving accuracy on ever larger datasets. Such large datasets pose a challenge to the current paradigms of data collection and data analysis at LHCb. Methods such as that described in Part III of this thesis, namely the use of smaller turbo data formats, will be necessary due to larger datasets and the increasing rates of data taking.

Appendix A

Time evolution of simulation samples

A.1 Time evolution of prompt simulation

The dependence of the prompt distribution of $\ln(\chi^2_{\text{IP}}(D^0))$ with decay time is investigated with a sample of prompt $D^0 \to K^-\pi^+$ simulation. The sample contains 1.2 million events that pass the trigger and stripping lines detailed in Chapter 5 and were generated using the conditions that they do not have a b-ancestor and have a decay length greater than 75 μm.

The $\ln(\chi^2_{\text{IP}})$ distribution of the sample is split into slices of decay time. In each slice a smeared $\ln(\chi^2_{\text{IP}})$ PDF, first introduced in Eq. (8.15), is fitted to the distribution. The value of the right tail parameter is fixed to its fit value from the first decay-time slice, where the statistics are highest. Example fits from three decay-time bins are shown in Fig. A.1. The poor quality of the fits is due to the large statistics at low decay times. In the main fit of Part II, two smeared $\ln(\chi^2_{\text{IP}})$ PDFs are used to describe signal. The variations of the PDF mean μ, and left tail parameter α_L, are shown in Fig. A.2. These values are used to motivate the starting parameters and configuration in the main fit of the A_Γ analysis of Part II.

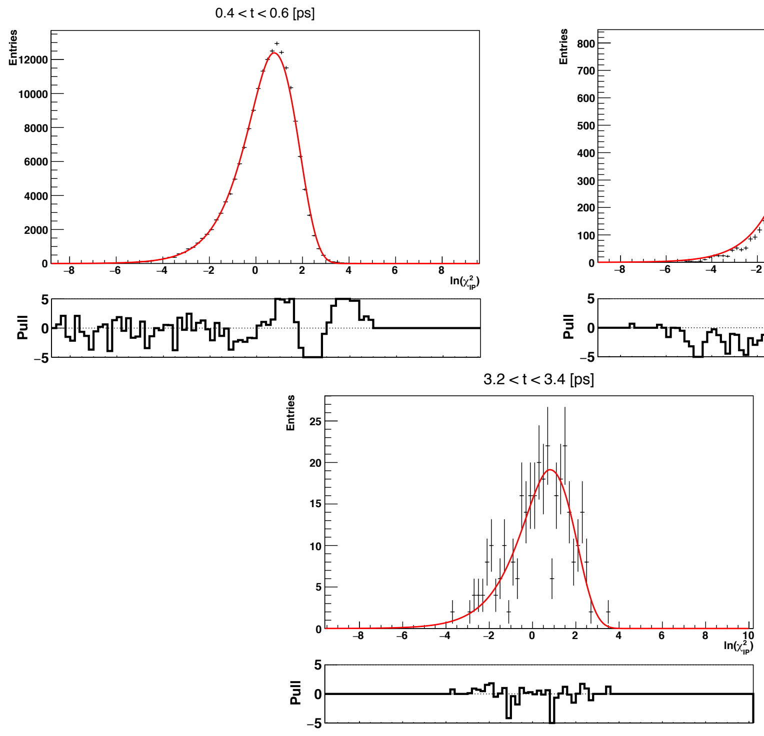


Figure A.1: Selection of smeared $\ln(\chi^2_{\rm IP})$ PDFs fit to prompt simulation in bins of decay-time.

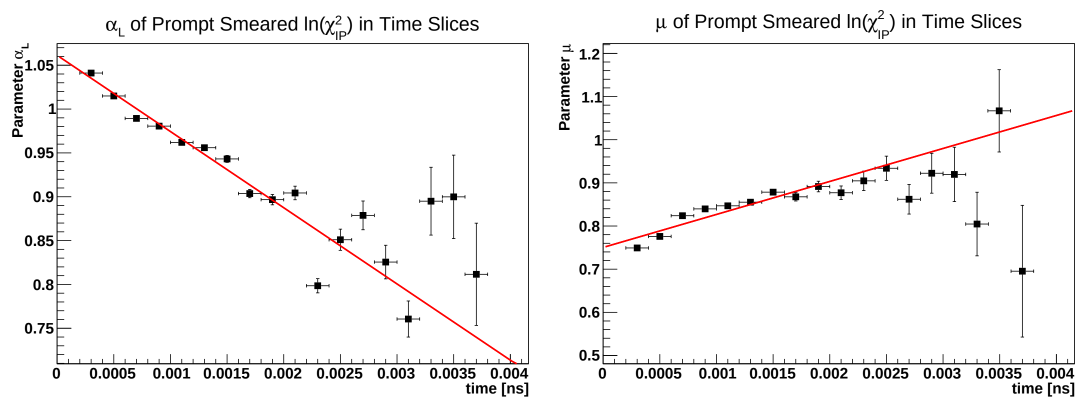


Figure A.2: Smeared $\ln(\chi^2_{\rm IP})$ left tail parameter (left) and mean (right) extracted from fits to prompt simulation in bins of decay-time.

A.2 Time evolution of secondary simulation

In this section the dependence of the secondary distribution of $\ln(\chi^2_{\mathrm{IP}})$ on decay time is investigated using a sample of $D^0 \to K^-\pi^+$ secondary simulation. This sample contains $570,000$ events that pass the trigger and stripping lines detailed in Chapter 5 and were generated using the conditions that they have a b-ancestor and a decay length greater than $75\,\mu\mathrm{m}$.

The $\ln(\chi^2_{\mathrm{IP}})$ distribution for the sample is split into slices of decay time. In each slice a smeared $\ln(\chi^2_{\mathrm{IP}})$ PDF is fitted to the distribution and the variation of the PDF mean μ, and tail parameters α_L and α_R are extracted. Example fits from three decay-time bins are shown in Fig. A.3. The μ, α_L and α_R parameters are fitted with an arctan function and two exponential functions, respectively, as shown in Figs. A.4 and A.5. The functions used to fit these parameters are used to model the decay-time dependence of $\ln(\chi^2_{\mathrm{IP}})$ in the main fit of the A_Γ analysis of Part II. The values of the parameters motivate the starting values used in the main fit of the A_Γ analysis.

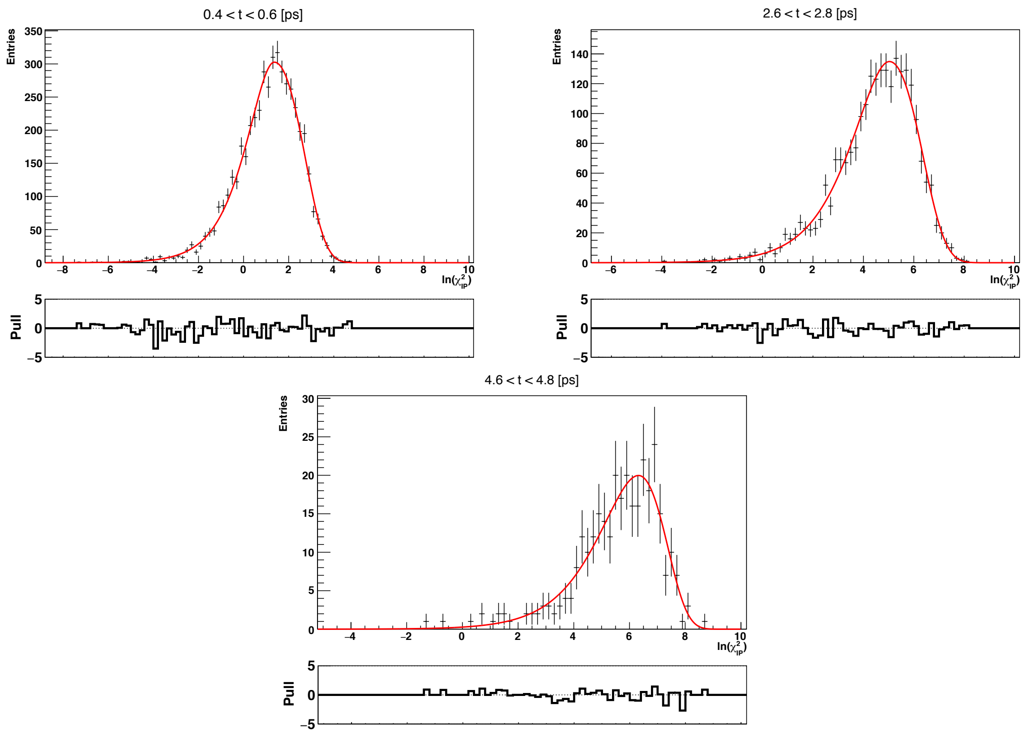


Figure A.3: Selection of smeared $\ln(\chi^2_{\mathrm{IP}})$ PDFs fit to secondary simulation in bins of decay time.

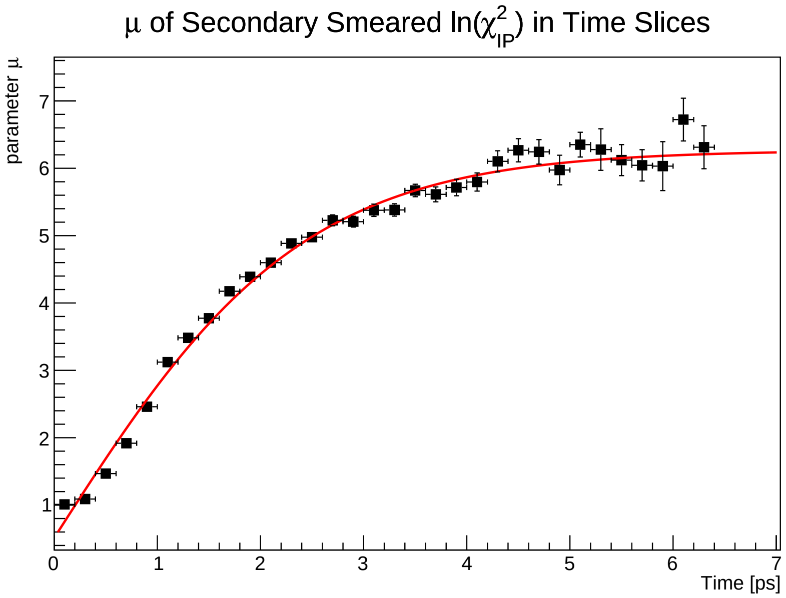


Figure A.4: Smeared $\ln(\chi^2_{\mathrm{IP}})$ mean parameter, μ, extracted from fits to secondary simulation in bins of decay time.

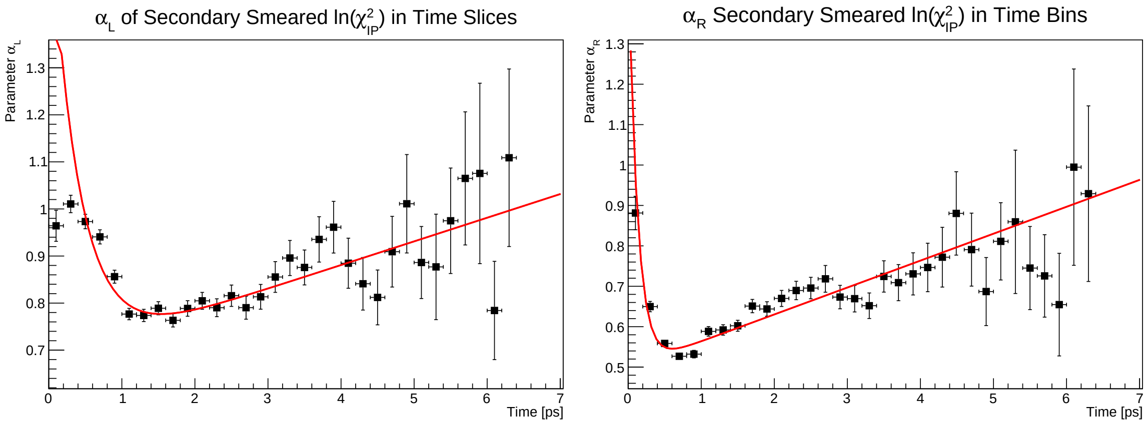


Figure A.5: Smeared $\ln(\chi^2_{\mathrm{IP}})$ distribution (left) left and (right) right tail parameters extracted from fits to secondary simulation in bins of decay time.

Appendix B

Fit parameter results

Shown here are tables of the final parameter values for the three final states of the A_Γ analysis and for both fit stages. The parameters are shown for the magnet polarity down, run block two sub-dataset, with D^0 flavour. The K^+K^- values can be found in Tables 7.2 and 8.1. The corresponding $\pi^+\pi^-$ results are in Tables B.1 and B.2. Finally the $K^-\pi^+$ results are shown in Tables B.3 and B.4.

Table B.1: Fit parameter results for the fit to $m(D^0)$ and Δm from one $D^0 \to \pi^+\pi^-$ sub-dataset with magnet polarity down and from the second run block. The symbols are explained in the text of Chapter 7. The parameter units are shown where relevant. In the case where a formula is presented in the first column, the value in the last column is the X variable of the formula. For example, $\sigma_{\Delta m_2,sig.} = (1.855)\ \sigma_{\Delta m_1,sig.}$.

Parameter	Fixed?	Value
$P(sig.)$	N	(0.5654 ± 0.0023)
$P(\pi_s)$	N	(0.2366 ± 0.0022)
$P(m_1\|sig.)$	N	(0.554 ± 0.019)
$\mu_{m,sig.}$	N	$(1866.963 \pm 0.020)\,\mathrm{MeV}/c^2$
$\sigma_{m_1,sig.}$	N	$(7.465 \pm 0.069)\,\mathrm{MeV}/c^2$
$\alpha_{sig.}$	Y	2.051
$n_{sig.}$	Y	1
$\sigma_{m_2,sig.}$	N	$(10.799 \pm 0.097)\,\mathrm{MeV}/c^2$
$e_{comb.}$	N	$(-0.00104 \pm 0.00019)[\,\mathrm{MeV}/c^2]^{-1}$
$P(\Delta m_1\|sig.)$	N	(0.165 ± 0.035)
$P(\Delta m_2\|sig.)$	N	(0.589 ± 0.024)
$\mu_{\Delta m_1,sig.}$	N	$(145.4220 \pm 0.0029)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_1,sig.}$	N	$(0.376 \pm 0.024)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_2,sig.} = \mathrm{X}\ \sigma_{\Delta m_1,sig.}$	N	(1.855 ± 0.069)
$\mu_{\Delta m_3,sig.}$	N	$(145.625 \pm 0.029)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_3,sig.} = \mathrm{X}\ \sigma_{\Delta m_2,sig.}$	N	(2.010 ± 0.056)
A_{π_s}	Y	$139.57\,\mathrm{MeV}/c^2$
B_{π_s}	N	(0.3688 ± 0.0077)
C_{π_s}	Y	$0\ [\mathrm{MeV}/c^2]^{-1}$
$A_{comb.}$	Y	$139.57\,\mathrm{MeV}/c^2$
$B_{comb.}$	N	(0.595 ± 0.029)
$C_{comb.}$	N	$(40.1 \pm 5.7)[\,\mathrm{MeV}/c^2]^{-1}$

Table B.2: Fit parameter results for the fit to D^0 decay-time and $\ln(\chi^2_{\mathrm{IP}})$ from one $D^0 \to \pi^+\pi^-$ sub-dataset with magnet polarity down and from the second run block. The symbols are explained in the text of Chapter 8. The parameter units are shown where relevant.

Parameter	Fixed?	Value
$P(prom.\|sig.)$	N	(0.92948 ± 0.00092)
τ	N	$(0.4141 \pm 0.0011)\,\mathrm{ps}$
$I_{\mu,prom.}$	N	(0.7271 ± 0.0074)
$S_{\mu,prom.}$	N	(0.2379 ± 0.0058)
$I_{\alpha_L,prom.}$	N	(1.0648 ± 0.0042)
$S_{\alpha_L,prom.}$	N	(-0.0759 ± 0.0034)
$\alpha_{R,prom.}$	N	(0.8821 ± 0.0060)
$P(sec.1\|sec.)$	N	(0.9556 ± 0.0029)
$\tau_{1,sec.}$	Y	$0.33\,\mathrm{ps}$
$\tau_{2,sec.}$	N	$(0.859 \pm 0.028)\,\mathrm{ps}$
$\tau_{3,sec.}$	N	$(0.2737 \pm 0.0099)\,\mathrm{ps}$
g_0	N	(-4.26 ± 0.11)
g_1	N	$(1.004 \pm 0.030)[\mathrm{ps}]^{-1}$
g_2	N	(10.00 ± 0.15)
$p_{L,0}$	N	(1.86 ± 0.20)
$p_{L,1}$	N	$(0.049 \pm 0.045)[\mathrm{ps}]^{-1}$
$p_{L,2}$	N	$(-0.82 \pm 0.32)[\mathrm{ps}]^{-1}$
$p_{R,0}$	N	(-0.656 ± 0.013)
$p_{R,1}$	N	$(0.0045 \pm 0.0055)[\mathrm{ps}]^{-1}$
$p_{R,2}$	Y	$-8\ [\mathrm{ps}]^{-1}$
$P(prom.\|\pi_s)$	N	(0.9156 ± 0.0018)
τ_{π_s}	N	$(0.4442 \pm 0.0021)\,\mathrm{ps}$

Table B.3: Fit parameter results for the fit to $m(D^0)$ and Δm from one $D^0 \to K^-\pi^+$ sub-dataset with magnet polarity down and from the second run block. The symbols are explained in the text of Chapter 7. The parameter units are shown where relevant. In the case where a formula is presented in the first column, the value in the last column is the X variable of the formula. For example $\sigma_{\Delta m_2,sig.} = (1.8870)\ \sigma_{\Delta m_1,sig.}$.

Parameter	Fixed?	Value
$P(sig.)$	N	(0.82283 ± 0.00036)
$P(\pi_s)$	N	(0.16882 ± 0.00037)
$P(m_1\|sig.)$	N	(0.3762 ± 0.0031)
$P(m_2\|sig.)$	N	(0.5664 ± 0.0033)
$\mu_{m,sig.}$	N	$(1866.8052 \pm 0.0034)\,\mathrm{MeV}/c^2$
$\sigma_{m_1,sig.}$	N	$(6.326 \pm 0.015)\,\mathrm{MeV}/c^2$
$\alpha_{sig.}$	Y	2.067
$n_{sig.}$	Y	1
$\sigma_{m_2,sig.}$	N	$(8.636 \pm 0.022)\,\mathrm{MeV}/c^2$
$\sigma_{m_3,sig.} = \mathrm{X}\ \sigma_{m_2,sig.}$	N	(1.941 ± 0.020)
$e_{comb.}$	N	$(-0.00260 \pm 0.00036)[\,\mathrm{MeV}/c^2]^{-1}$
$P(\Delta m_1\|sig.)$	N	(0.1946 ± 0.0057)
$P(\Delta m_2\|sig.)$	N	(0.5909 ± 0.0039)
$\mu_{\Delta m_1,sig.}$	N	$(145.42107 \pm 0.00050)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_1,sig.}$	N	$(0.3843 \pm 0.0035)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_2,sig.} = \mathrm{X}\ \sigma_{\Delta m_1,sig.}$	N	(1.8870 ± 0.0099)
$\mu_{\Delta m_2,sig.}$	N	$(145.6557 \pm 0.0054)\,\mathrm{MeV}/c^2$
$\sigma_{\Delta m_3,sig.} = \mathrm{X}\ \sigma_{\Delta m_2,sig.}$	N	(1.9918 ± 0.0084)
A_{π_s}	Y	$139.57\,\mathrm{MeV}/c^2$
B_{π_s}	N	(0.3641 ± 0.0018)
C_{π_s}	Y	$0\ [\mathrm{MeV}/c^2]^{-1}$
$A_{comb.}$	Y	$139.57\ \ \mathrm{MeV}/c^2$
$B_{comb.}$	N	(-0.341 ± 0.013)
$C_{comb.}$	Y	$0\ [\mathrm{MeV}/c^2]^{-1}$

Table B.4: Fit parameter results for the fit to D^0 decay time and $\ln(\chi^2_{\text{IP}})$ from one $D^0 \to K^-\pi^+$ sub-dataset with magnet polarity down and from the second run block. The symbols are explained in the text of Chapter 8. The parameter units are shown where relevant.

Parameter	Fixed?	Value
$P(prom.\|sig.)$	N	(0.92051 ± 0.00017)
τ	N	$(0.41534 \pm 0.00021)\,\text{ps}$
$P(\chi 1\|prom.)$	N	(0.98747 ± 0.00044)
$I_{\mu,1,prom.}$	N	(0.7876 ± 0.0016)
$S_{\mu,prom.}$	N	(0.1741 ± 0.0012)
$I_{\alpha_L,prom.}$	N	(1.05777 ± 0.00094)
$S_{\alpha_L,prom.}$	N	(-0.06498 ± 0.00077)
$\alpha_{R,1,prom.}$	N	(0.8894 ± 0.0017)
$I_{\mu,2,prom.}$	N	(2.892 ± 0.023)
$\alpha_{L,2,prom.}$	Y	0.9622
$\alpha_{R,2,prom.}$	N	(2.099 ± 0.064)
$P(sec.1\|sec.)$	N	(0.91073 ± 0.00096)
$\tau_{1,sec.}$	Y	$0.33\,\text{ps}$
$\tau_{2,sec.}$	N	$(0.8582 \pm 0.0035)\,\text{ps}$
$\tau_{3,sec.}$	N	$(0.2736 \pm 0.0020)\,\text{ps}$
g_0	N	(-4.220 ± 0.023)
g_1	N	$(1.0176 \pm 0.0064)[\text{ps}]^{-1}$
g_2	N	(9.714 ± 0.030)
$p_{L,0}$	N	(1.397 ± 0.028)
$p_{L,1}$	N	$(0.0605 \pm 0.0069)[\text{ps}]^{-1}$
$p_{L,2}$	N	$(-0.921 \pm 0.052)[\text{ps}]^{-1}$
$p_{R,0}$	N	(-0.6422 ± 0.0023)
$p_{R,1}$	N	$(0.01700 \pm 0.00099)[\text{ps}]^{-1}$
$p_{R,2}$	Y	$-8\ [\text{ps}]^{-1}$
$P(prom.\|\pi_s)$	N	(0.92340 ± 0.00040)
τ_{π_s}	N	(0.42018 ± 0.00056)

Appendix C

Fits to Δm for the $\tau(D^0)$ analysis

The Δm fits in all eight decay-time bins, and for both magnet polarities of the $\tau(D^0)$ analysis are shown here. The magnet polarity up and magnet polarity down fits of the main fit method are shown in Figs. C.1 and C.2, respectively.

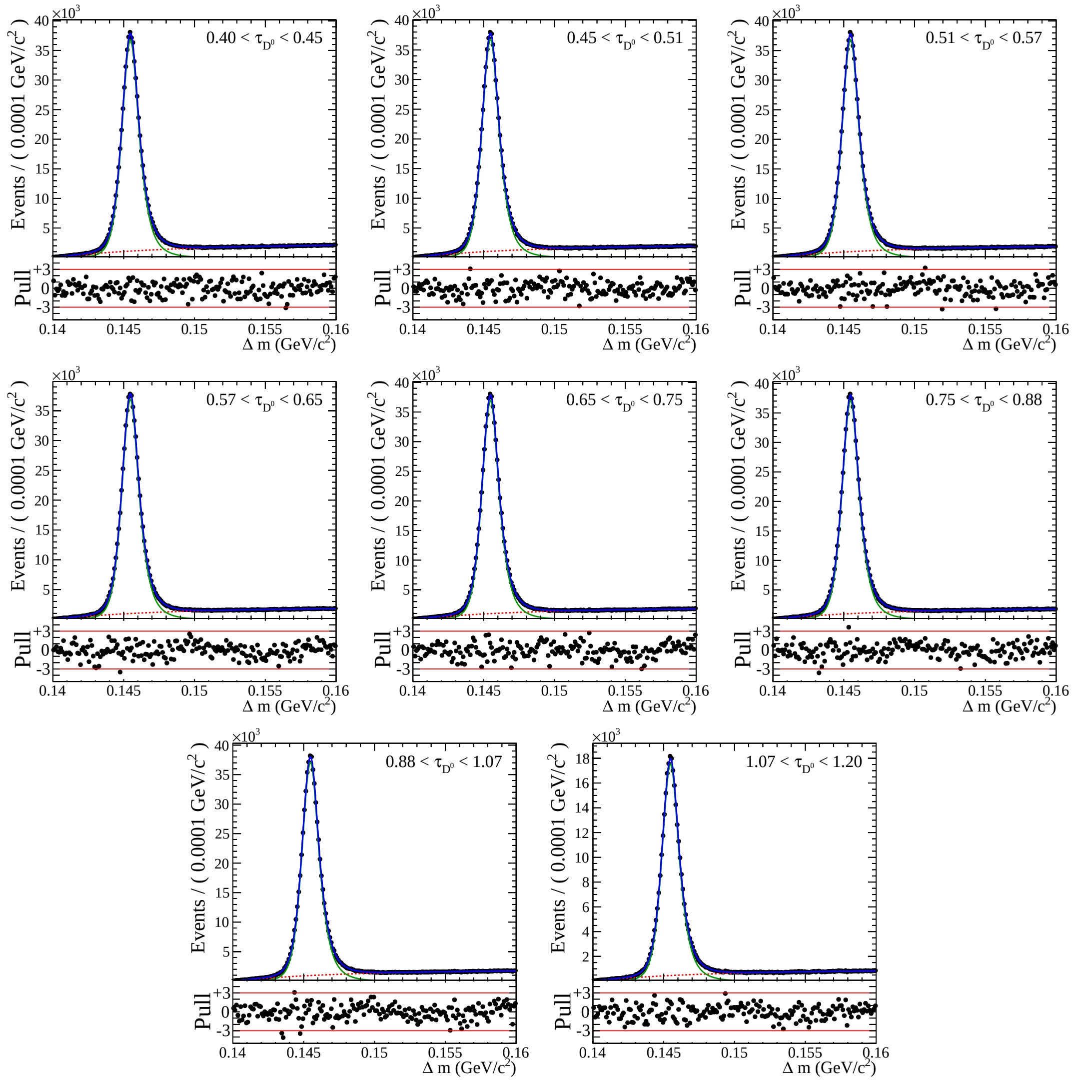


Figure C.1: Fits to the Δm distributions in all decay-time bins and for the magnet polarity up sub-dataset. From these fits the number of prompt signal events in each bin is extracted. The fit components are (blue) the total fit, (green) the prompt signal component and (dashed red) the remaining backgrounds. The bottom frame in each plot shows the pull distribution, or the difference between the total fit and the data, divided by the data uncertainty. The pull distributions show that the agreement between the fit and the data is excellent.

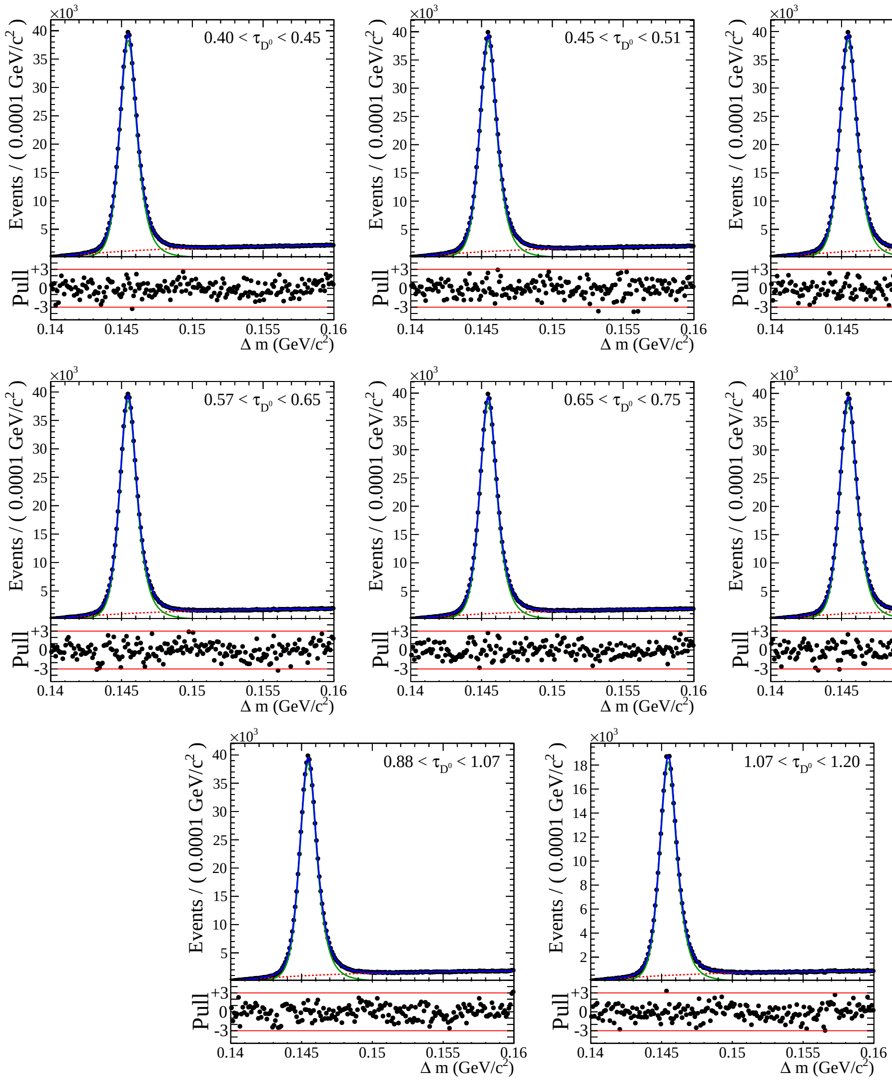


Figure C.2: Fits to the Δm distributions in all decay-time bins and for the magnet polarity down sub-dataset. From these fits the number of prompt signal events in each bin is extracted. The fit components are (blue) the total fit, (green) the prompt signal component and (dashed red) the remaining backgrounds. The bottom frame in each plot shows the pull distribution, or the difference between the total fit and the data, divided by the data uncertainty. The pull distributions show that the agreement between the fit and the data is excellent.

Appendix D

Cut based separation with decay chain isolation

In this appendix, a selection requirement is applied to one of the isolation observables of Section 15.5 to separate prompt signal and secondary background events. The requirement is applied to the single isolation observable that gives the best discrimination between prompt and secondary simulation. The two-track mass for the D^0, m_2, gives the best separation as shown in Fig. 15.5 (top left). Shown in Fig. D.1 are the 2D plots of this variable against $\ln(\chi^2_{\text{IP}})$, where the shape of the prompt and secondary distributions are seen to be significantly different.

To determine the optimum selection requirement on $\ln(m_2)$, the metric $S/\sqrt{B}$ is used, where S is the number of prompt signal events and B is the number of background secondary events. The relative fraction of prompt and secondary

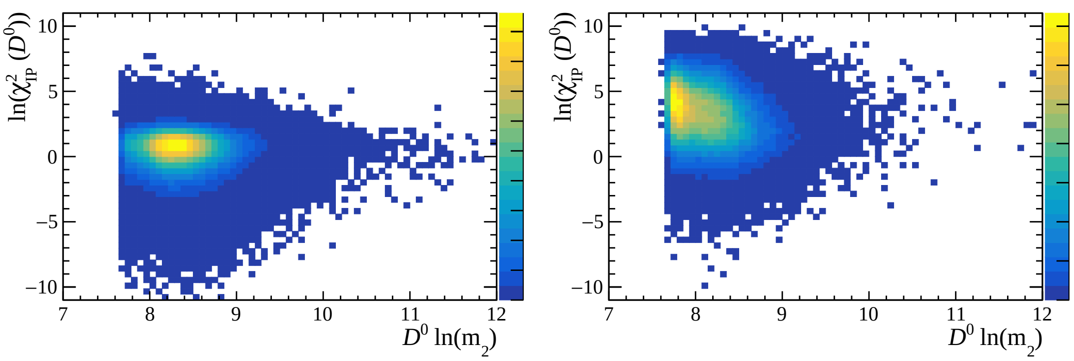
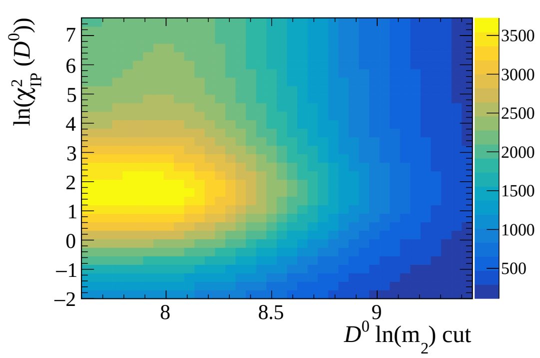


Figure D.1: Normalised distributions of prompt (left) and secondary (right) simulation for the natural logarithm of the two-track mass at the D^0 vertex against $\ln(\chi^2_{\text{IP}})$. The differences in the position of the peaks and in their shapes are evident.



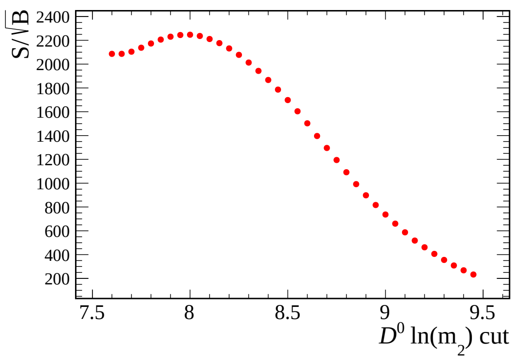


Figure D.2: Distributions of the prompt and secondary separation metric $S/\sqrt{B}$ for different $\ln(\chi^2_{\rm IP})$ and $\ln(m_2)$ requirement values. Shown are (left) the metric with two cuts, one on $\ln(\chi^2_{\rm IP})$ and one on $\ln(m_2)$, and (right) the metric for cuts on $\ln(m_2)$ only. The prompt and secondary simulation samples are normalised such that the secondary is 10% of the sum of both samples before applying these cuts.

events cannot be determined from the two separate simulation samples. Thus, to compare the prompt and secondary simulation samples a global secondary fraction of 10% must be assumed. This is taken from a comparison of the $\ln(\chi^2_{\rm IP})$ distributions of the simulation samples with the 2016 data sample of Part III, and is similar to the secondary fraction observed in the A_Γ measurement ($\sim 7\%$). The prompt and secondary samples are scaled such that the sum of both samples (100%) comprises of 90% prompt events and 10% secondary events. Fig. D.2 shows the metric against selection requirements on $\ln(m_2)$ and for a pair of requirements on $\ln(m_2)$ and $\ln(\chi^2_{\rm IP})$. The $\ln(m_2)$ cut is not as powerful as a cut on $\ln(\chi^2_{\rm IP})$, as its optimal lone cut at a value of eight only reduces the secondary fraction to 7%. The best lone cut for $\ln(\chi^2_{\rm IP})$ at 1.6 reduces the secondary fraction to about 3%. However, both cuts remove a different tail of the secondary distribution, so applying the cuts together can give a better performance. The optimum cuts from Fig. D.2 (left) are $\ln(\chi^2_{\rm IP}) < 1.6$ and $\ln(m_2) > 7.9$, although a cut of $\ln(m_2) > 8.1$ has a similar significance but removes more secondary. With both cuts applied the secondary is reduced to 2.5% of the total simulation sample, assuming it was 10% of the initial sample.

Bibliography

[1] J. C. Maxwell, *"VIII. A dynamical theory of the electromagnetic field"*, Phil. Trans. R. Soc. A **155** (1865) 459.

[2] C. D. Anderson, *"The positive electron"*, Phys. Rev. **43** (1933) 491.

[3] ACME Collaboration, J. Baron *et al.*, *"Order of magnitude smaller limit on the electric dipole moment of the electron"*, Science **343** (2014) 269.

[4] Super-Kamiokande Collaboration, Y. Fukuda *et al.*, *"Evidence for oscillation of atmospheric neutrinos"*, Phys. Rev. Lett. **81** (1998) 1562.

[5] MINOS Collaboration, D. G. Michael *et al.*, *"Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam"*, Phys. Rev. Lett. **97** (2006) 191801.

[6] Heavy flavour averaging group (HFLAV), Y. Amhis *et al.*, *"Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016"*, `arXiv:1612.07233`.

[7] LHCb Collaboration, R. Aaij *et al.*, *"Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays"*, JHEP **08** (2017) 055.

[8] LHCb Collaboration, R. Aaij *et al.*, *"Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays"*, Phys. Rev. Lett. **113** (2014) 151601.

[9] Particle Data Group, C. Patrignani *et al.*, *"Review of particle physics, 2016-2017"*, Chin. Phys. C **40** (2016), no. 10.

[10] M. Gell-Mann, *"A schematic model of baryons and mesons"*, Phys. Lett. **8** (1964) 214.

[11] V. E. Barnes *et al.*, *"Observation of a hyperon with strangeness -3"*, Phys. Rev. Lett. **12** (1964) 204.

[12] O. W. Greenberg, *"Spin and unitary spin independence in a paraquark model of baryons and mesons"*, Phys. Rev. Lett. **13** (1964) 598.

[13] C. S. Wu *et al.*, *"Experimental test of parity conservation in beta decay"*, Phys. Rev. **105** (1957) 1413.

[14] A. Banfi *et al.*, *"Lecture notes for the 2015 HEP school for experimental high energy physics students"*, RAL-TR-2015-005, STFC, September 2015.

[15] M. Peskin and D. Schroeder, *"An introduction to quantum field theory"*, Westview Press, 1995.

[16] F. Halzen and D. Martin, *"Quarks and leptons: An introductory course in modern particle physics"*, John Wiley and Sons, 1984.

[17] F. Englert and R. Brout, *"Broken symmetry and the mass of gauge vector mesons"*, Phys. Rev. Lett. **13** (1964) 321.

[18] P. W. Higgs, *"Broken symmetries and the masses of gauge bosons"*, Phys. Rev. Lett. **13** (1964) 508.

[19] N. Cabibbo, *"Unitary symmetry and leptonic decays"*, Phys. Rev. Lett. **10** (1963) 531.

[20] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *"Evidence for the 2π decay of the K_2^0 meson"*, Phys. Rev. Lett. **13** (1964) 138.

[21] M. Kobayashi and T. Maskawa, *"CP violation in the renormalizable theory of weak interaction"*, Prog. Theor. Phys. **49** (1973) 652.

[22] M. S. Sozzi, *"Discrete symmetries and CP violation"*, Oxford University Press, 2008.

[23] S. L. Glashow, J. Iliopoulos, and L. Maiani, *"Weak interactions with lepton-hadron symmetry"*, Phys. Rev. **D2** (1970) 1285.

[24] S. Reichert, *"Measurement of the mixing parameters of neutral charm mesons and search for indirect CP violation with $D^0 \to K_S^0\pi^+\pi^-$ decays at LHCb"*, Dec, 2015. CERN-THESIS-2015-348.

[25] Y. Nir, *"CP violation in meson decays"*, in *Proceedings, 3rd Latin American CERN-CLAF School, Malargue, Argentina*, 2005. `arXiv:hep-ph/0510413`.

[26] A. L. Kagan and M. D. Sokoloff, *"On indirect CP violation and implications for D^0-$\overline{D}^0$ and B_s^0-$\overline{B}_s^0$ mixing"*, Phys. Rev. **D80** (2009) 076008.

[27] BaBar Collaboration, B. Aubert *et al.*, *"Observation of direct CP violation in $B^0 \to K^+\pi^-$ decays"*, Phys. Rev. Lett. **93** (2004) 131801.

[28] LHCb Collaboration, R. Aaij *et al.*, *"First observation of CP violation in the decays of B_s^0 mesons"*, Phys. Rev. Lett. **110** (2013), no. 22 221601.

[29] M. Gersabeck *et al.*, *"On the interplay of direct and indirect CP violation in the charm sector"*, J. Phys. **G39** (2012) 045005.

[30] Belle Collaboration, M. Staric, *"New Belle results on D^0-$\overline{D}^0$ mixing"*, `arXiv:1212.3478`.

[31] BaBar Collaboration, J. P. Lees *et al.*, *"Measurement of D^0-$\overline{D}^0$ mixing and CP violation in two-body D^0 decays"*, Phys. Rev. **D87** (2013) 012004.

[32] LHCb Collaboration, R. Aaij *et al.*, *"Measurement of mixing and CP violation parameters in two-body charm decays"*, JHEP **04** (2012) 129.

[33] LHCb Collaboration, R. Aaij *et al.*, *"Measurements of indirect CP asymmetries in $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ decays"*, Phys. Rev. Lett. **112** (2014) 041801.

[34] CDF Collaboration, T. A. Aaltonen *et al.*, *"Measurement of indirect CP-violating asymmetries in $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ decays at CDF"*, Phys. Rev. **D90** (2014), no. 11 111103.

[35] LHCb Collaboration, R. Aaij *et al.*, *"Measurement of indirect CP asymmetries in $D^0 \to K^-K^+$ and $D^0 \to \pi^-\pi^+$ decays using semileptonic B decays"*, JHEP **04** (2015) 043.

[36] L. Evans and P. Bryant, *"LHC machine"*, JINST **3** (2008), no. 08 S08001.

[37] O. S. Brüning *et al.*, *"LHC design report"*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004.

[38] ATLAS Collaboration, G. Aad *et al.*, *"The ATLAS Experiment at the CERN Large Hadron Collider"*, JINST **3** (2008) S08003.

[39] CMS Collaboration, S. Chatrchyan *et al.*, *"The CMS Experiment at the CERN LHC"*, JINST **3** (2008) S08004.

[40] ALICE Collaboration, K. Aamodt *et al.*, *"The ALICE experiment at the CERN LHC"*, JINST **3** (2008) S08002.

[41] MoEDAL Collaboration, J. Pinfold *et al.*, *"Technical design report of the MoEDAL experiment"*, Tech. Rep. CERN-LHCC-2009-006. MoEDAL-TDR-001, Jun, 2009.

[42] LHCf Collaboration, O. Adriani *et al.*, *"LHCf experiment: Technical design report"*, CERN, Geneva, 2006.

[43] TOTEM Collaboration, W. Kienzle *et al.*, *"TOTEM, total cross section, elastic scattering and diffraction dissociation at the LHC: Technical proposal"*, Tech. Rep. CERN-LHCC-99-007. LHCC-P-5, CERN, Geneva, Mar, 1999.

[44] C. Lefèvre, *"The CERN accelerator complex. Complexe des accélérateurs du CERN"*, Dec, 2008.

[45] A. D. Rosso, *"Particle kickers. Aiguiller les particules"*, Jun, 2014.

[46] LHCb Collaboration, A. Alves *et al.*, *"The LHCb detector at the LHC"*, JINST **3** (2008), no. 08 S08005.

[47] LHCb collaboration, R. Aaij *et al.*, *"LHCb detector performance"*, Int. J. Mod. Phys. **A30** (2015) 1530022.

[48] LHCb Collaboration, R. Aaij *et al.*, *"Precision luminosity measurements at LHCb"*, JINST **9** (2014), no. 12 P12005.

[49] LHCb collaboration, R. Aaij *et al.*, *"Performance of the LHCb Vertex Locator"*, JINST **9** (2014) P09007.

[50] R. Arink *et al.*, *"Performance of the LHCb Outer Tracker"*, JINST **9** (2014) P01002.

[51] B. R. Martin and G. Shaw, *"Particle physics"*, The Manchester Physics Series, WILEY, 2008.

[52] M. Adinolfi *et al.*, *"Performance of the LHCb RICH detector at the LHC"*, Eur. Phys. J. **C73** (2013) .

[53] R. W. Forty and O. Schneider, *"RICH pattern recognition"*, Tech. Rep. LHCb-98-040, CERN, Geneva, Apr, 1998.

[54] LHCb Collaboration, S. Amato *et al.*, *"LHCb calorimeters: Technical design report"*, Technical Design Report LHCb, CERN, Geneva, 2000.

[55] I. Machikhiliyan, *"The LHCb electromagnetic calorimeter"*, Tech. Rep. LHCb-PROC-2009-005, LHCb-CONF-2009-005, CERN, Geneva, 2009.

[56] CMS and LHCb Collaborations, Khachatryan *et al.*, *"Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data"*, Nature **522** (2014) 68.

[57] A. Puig, *"The LHCb trigger in 2011 and 2012"*, Tech. Rep. LHCb-PUB-2014-046, CERN-LHCb-PUB-2014-046, CERN, Geneva, 2014.

[58] S. Benson *et al.*, *"The LHCb turbo stream"*, J. Phys. Conf. Ser. **664** (2015) .

[59] S. Miglioranzi *et al.*, *"The LHCb simulation application, Gauss: design, evolution and experience"*, Tech. Rep. LHCb-PROC-2011-006, CERN-LHCb-PROC-2011-006, CERN, Geneva, Jan, 2011.

[60] T. Sjostrand, S. Mrenna, and P. Z. Skands, *"A brief introduction to PYTHIA 8.1"*, Comput. Phys. Commun. **178** (2008) 852.

[61] A. Buckley *et al.*, *"LHAPDF6: parton density access in the LHC precision era"*, Eur. Phys. J. **C75** (2015) 132.

[62] D. J. Lange, *"The EvtGen particle decay simulation package"*, Nucl. Instrum. Meth. **A462** (2001) 152.

[63] J. Allison *et al.*, *"Geant4 developments and applications"*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[64] LHCb collaboration, R. Aaij *et al.*, *"Measurement of mixing and CP violation parameters in two-body charm decays"*, JHEP **04** (2012) 129.

[65] LHCb collaboration, R. Aaij *et al.*, *"Measurements of indirect CP asymmetries in $D^0 \to K^-K^+$ and $D^0 \to \pi^-\pi^+$ decays"*, Phys. Rev. Lett. **112** (2014) 041801.

[66] LHCb Collaboration, R. Aaij *et al.*, *"Measurement of the CP violation parameter A_Γ in $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ decays"*, Phys. Rev. Lett. **118** (2017) 261803.

[67] P. Marino, *"Measurement of the CP violation parameter A_Γ in $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ decays"*, 2017. CERN-THESIS-2017-007.

[68] S. Tolk *et al.*, *"Data driven trigger efficiency determination at LHCb"*, Tech. Rep. LHCb-PUB-2014-039, CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014.

[69] V. V. Gligorov, *"A single track HLT1 trigger"*, Tech. Rep. LHCb-PUB-2011-003, CERN-LHCb-PUB-2011-003, LHCb-INT-2010-053, CERN, Geneva, Jan, 2011.

[70] O. Behnke *et al.*, *"Data analysis in high energy physics: A practical guide to statistical methods"*, Wiley-VCH, 2013.

[71] F. James and M. Roos, *"Minuit: A system for function minimization and analysis of the parameter errors and correlations"*, Comput. Phys. Commun. **10** (1975) 343.

[72] M. Pivk and F. R. Le Diberder, *"SPlot: A statistical tool to unfold data distributions"*, Nucl. Instrum. Meth. **A555** (2005) 356.

[73] D. Scott, *"Multivariate density estimation: Theory, practice, and visualization"*, John Wiley and Sons Inc, 1992.

[74] J. E. Gaiser, *"Charmonium spectroscopy from radiative decays of the J/ψ and ψ'"*, PhD thesis, SLAC, 1982.

[75] M. Smith, *"Measurement of indirect CP violation in charm at LHCb"*, Sep, 2015. CERN-THESIS-2015-377.

[76] L. Anderlini *et al.*, *"The PIDCalib package"*, Tech. Rep. LHCb-PUB-2016-021, CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.

[77] J. Rademacker, *"Reduction of statistical power per event due to upper lifetime cuts in lifetime measurements"*, Nucl. Instrum. Meth. **A570** (2007) 525.

[78] CDF Collaboration, T. Aaltonen *et al.*, *"Measurement of the B^- lifetime using a simulation free approach for trigger bias correction"*, Phys. Rev. **D83** (2011) 032008.

[79] V. V. Gligorov, *"Measurement of the CKM angle γ and B meson lifetimes at the LHCb detector"*, PhD thesis, 2008, CERN-THESIS-2008-044.

[80] M. Gersabeck, *"Alignment of the LHCb vertex locator and lifetime measurements of two-body hadronic final states"*, PhD thesis, 2009, CERN-THESIS-2009-118.

[81] Y. Amhis *et al.*, *"b-hadrons absolute lifetime measurements"*, Feb, 2014. LHCb-ANA-2011-055. Linked to LHCb-PAPER-2013-065.

[82] LHCb Collaboration, R. Aaij *et al.*, *"Selections and lifetime measurements for exclusive $b \to J/\psi X$ decays with $J/\psi \to \mu^+\mu^-$ with 2010 data"*, LHCb-ANA-2011-001.

[83] L. Lyons, D. Gibaut, and P. Clifford, *"How to combine correlated estimates of a single physical quantity"*, Nucl. Instrum. Meth. **A270** (1988) 110.

[84] FOCUS Collaboration, J. M. Link *et al.*, *"New measurements of the D^0 and D^+ lifetimes"*, Phys. Lett. **B537** (2002) 192.

[85] SELEX Collaboration, A. Kushnirenko *et al.*, *"Precision measurements of the Λ_c^+ and D^0 lifetimes"*, Phys. Rev. Lett. **86** (2001) 5243.

[86] E791 Collaboration, E. M. Aitala *et al.*, *"Measurements of lifetimes and a limit on the lifetime difference in the neutral D meson system"*, Phys. Rev. Lett. **83** (1999) 32.

[87] CLEO Collaboration, G. Bonvicini *et al.*, *"Measurement of charm meson lifetimes"*, Phys. Rev. Lett. **82** (1999) 4586.

[88] E687 Collaboration, P. L. Frabetti *et al.*, *"Precise measurements of the D^0 and D^+ meson lifetimes"*, Phys. Lett. **B323** (1994) 459.

[89] M. W. Kenzie and V. Gligorov, *"Lifetime unbiased beauty and charm triggers at LHCb"*, Tech. Rep. LHCb-PUB-2015-026, CERN-LHCb-PUB-2015-026, CERN, Geneva, Oct, 2016.

[90] Y. Freund, *"Boosting a weak learning algorithm by majority"*, Inf. Comput. **121** (1995) 256.

[91] R. E. Schapire and Y. Freund, *"A decision-theoretic generalization of on-line learning and an application to boosting"*, J. Comput. Syst. Sci. **55** (1997) 119.

[92] J. H. Friedman, *"Greedy function approximation: A gradient boosting machine"*, Annals of Statistics **29** (2000).

[93] T. Hastie, R. Tibshirani, and J. Friedman, *"The elements of statistical learning"*, Springer, 2001.

[94] A. Rogozhnikov *et al.*, *"New approaches for boosting to uniformity"*, JINST **10** (2015), no. 03 T03002.

[95] J. Stevens and M. Williams, *"uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers"*, JINST **8** (2013), no. 12 P12013.

[96] W. Hulsbergen *et al.*, *"Decay chain fitting with a Kalman filter"*, Nucl. Instrum. Meth. **A552** (2005) , `arXiv:physics/0503191`.

[97] A. Schwartz, *"Prospects on time-dependent CPV and mixing measurements at Belle II"*, 2016. 9th International Workshop on the CKM unitarity triangle.